

## PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: **C12N 15/10, 15/62, A01H 5/00**

(11) International Publication Number: **WO 00/61740**

(43) International Publication Date: **19 October 2000 (19.10.00)**

(21) International Application Number: **PCT/US00/09283**

(22) International Filing Date: **6 April 2000 (06.04.00)**

(30) Priority Date: **10 April 1999 (10.04.99)**

US

(71) Applicant (for all designated States except US): **MAXYGEN, INC. (USUS); 515 Galveston Drive, Redwood City, CA 94063 (US).**

(72) Inventors and

(73) Inventors/Applicants (for US only): **YUAN, Ling (USUS); 4428 Century Chan Drive, El Rancho, CA 95016 (US); RALLABADI, Sun, A1 (CNUS); 864 Timothy Drive, Mountain View, CA 94043 (US); LASSITER, Michael (USUS); 721 Falcon Avenue, Davis, CA 95616 (US).**

(74) Agents: **QUNE, Jonathan, Alan; The Law Offices of Jonathan Alan Qune, P.O. Box 438, Alameda, CA 94601 (US) et al.**

Published  
With international search report  
before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: **MODIFIED LIPID PRODUCTION**

(57) Abstract

Methods of modifying lipid production in cells and whole organisms are provided. Single genes, operons, lipid biosynthetic cycles and whole genomes can be recombined to produce cells and organisms with desirable lipid synthetic or metabolic activity. Libraries of recombinant lipid synthetic nucleic acids and organisms are also provided.

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SE	Sweden
BA	Bosnia and Herzegovina	GB	United Kingdom	MD	Moldova	SG	Singapore
BB	Barbados	GE	Georgia	MG	Madagascar	TH	Thailand
BE	Belgium	GR	Greece	ML	Mali	TI	Taiwan
BF	Burkina Faso	HN	Honduras	MR	Morocco	TR	Turkey
BG	Bulgaria	IE	Ireland	MT	Malta	TT	Trinidad and Tobago
BH	Bahrain	IL	Israel	MY	Malaysia	UA	Ukraine
BR	Brazil	IN	India	NP	Nepal	UG	Uganda
BT	Bhutan	JP	Japan	NI	Nicaragua	US	United States of America
CA	Canada	KE	Kenya	NL	Netherlands	VN	Viet Nam
CC	Cocos (Keeling) Islands	KG	Kyrgyzstan	NO	Norway	VU	Vanuatu
CD	Congo	KR	Republic of Korea	NZ	New Zealand	ZW	Zimbabwe
CE	Czech Republic	KZ	Kazakhstan	PE	Peru		
CH	Switzerland	LA	Laos	PL	Poland		
CI	Cote d'Ivoire	LB	Lebanon	PT	Portugal		
CN	China	LC	Liechtenstein	RO	Romania		
CO	Colombia	LI	Liechtenstein	SD	Sudan		
CR	Costa Rica	LK	Sri Lanka	SG	Singapore		
DE	Germany						
DK	Denmark						
EE	Estonia						

## FOR THE PURPOSES OF INFORMATION ONLY

# MODIFIED LIPID PRODUCTION

## CROSS REFERENCE TO RELATED APPLICATIONS

This application is a non-provisional regular utility application filing based on provisional application USSN 60/128,707, filed April 10, 1999, entitled "MODIFIED LIPID PRODUCTION" by Yuan and Raillard. The present application claims priority to and benefit of this earlier provisional application, pursuant to 37 C.F.R. 1.119(e) and pursuant to any other applicable U.S. statute or rule.

## FIELD OF THE INVENTION

This invention relates to the application of DNA shuffling technologies to nucleic acids coding factors which affect lipid biosynthesis and metabolism. The invention provides strategies for modifying every relevant lipid synthetic gene, coding sequence or promoter, as well as all other relevant DNA sequences involved in fatty acid biosynthetic pathways.

## BACKGROUND OF THE INVENTION

Fatty acids are organic acids having a hydrocarbon chain of about 4 to 24 carbons. Many different fatty acids are known which differ from each other in chain length and in the presence, number and position of double or triple bonds.

In cells, fatty acids typically exist in covalently bound forms, with the carboxyl portion of the fatty acid being referred to as a fatty acyl group. The chain length and degree of saturation of these molecules is often depicted by the formula  $CX:Y$ , where "X" indicates number of carbons and "Y" indicates number of double bonds (e.g. oleate, an 18 carbon molecule with 1 double bond, is shown as  $C18:1$ , or more simply, 18:1).

Fatty acids are the major components of all edible and industrial oils. The fatty acid composition of an oil determines its physical and chemical properties and thus its uses. For example, the fatty acyl group in a lipid molecule can be covalently bound to another group via a different linkage. It can be linked, e.g., through a thioester bond to an acyl carrier protein (ACP) or a Coenzyme A (CoA) to form an acyl-ACP or acyl-CoA, respectively. Similarly, it can also be linked through an ester bond to a fatty

alcohol to form a wax, and three acyl groups can be linked to a glycerol molecule to form triacylglycerol (triglyceride).

There are approximately 40 enzymes known to be involved in fatty acid biosynthesis. In plants, genetic engineering has provided for the generation of novel seed oils which have fatty acid compositions different from naturally existing plant oils (e.g., canola, containing laurate). Novel fatty acids can also be produced as a result of the expression of a DNA sequence foreign to other organisms. For example, fungi can be transformed with recombinant DNA and used in fermentation for the production of certain fatty acids. However, major hurdles to novel fatty acid production exist, due, e.g., to the difficulty of isolating enzymes with desired specificity, poor kinetics, different gene codon usage, incompatibility with the substrate pool and lack of strong promoters with appropriate expression timing and desirable tissue specificity. It would be beneficial to modify the properties of the enzymes and factors involved in fatty acid biosynthesis and metabolism to overcome these hurdles.

Although most of the genes encoding these enzymes are cloned and sequenced, achieving this goal is by no means trivial. The understanding of these enzymes and their genes at the biochemical and molecular level is poor.

Surprisingly, the present invention provides a strategy for solving each of the problems outlined above, as well as providing a variety of other features which will become apparent upon complete review of the following.

## SUMMARY OF THE INVENTION

The invention provides methods of modifying any or all genes encoding lipid synthesis-related enzymes and all nucleic acid elements involved in controlling the expressions of these genes and the cellular localization of the gene products. In addition, this invention also covers the use of these modified DNA sequences in transgenic (recombinant) organisms, including bacteria, fungi, algae, plants and animals. Modification of the fatty acid composition of a transgenic organism is achieved as a result of the introduction of DNA sequence(s) which have been modified, e.g., through DNA shuffling.

Accordingly, the invention provides a method of making a nucleic acid encoding a lipid biosynthetic activity. In the method, a plurality of parental nucleic acids are recombined to produce one or more recombinant lipid biosynthetic nucleic acids comprising distinct or improved lipid biosynthetic activities. The one or more

recombinant lipid biosynthetic nucleic acids are selected for one or more encoded lipid biosynthetic activities or, e.g., for reduced or enhanced encoded polypeptide expression or stability. These selection steps provide a selected shuffled lipid biosynthetic nucleic acid which encodes one or more selected lipid biosynthetic activities.

5 A variety of lipid biosynthetic activities can be selected, separately or in combination, including: modulation of lipid saturation for one or more selected lipids produced by a lipid synthetic pathway comprising activity encoded by the one or more selected shuffled lipid biosynthetic nucleic acids, modulation of fatty acid composition in a transgenic plant, algae, animal, bacteria, fungus or other organism expressing the selected shuffled lipid biosynthetic nucleic acid, modulation of fatty alcohol composition in a transgenic plant, algae, animal, bacteria, fungus or other organism expressing the selected shuffled lipid biosynthetic nucleic acid, modulation of a wax composition in a transgenic plant, algae, animal, bacteria, fungus or other organism expressing the selected shuffled lipid biosynthetic nucleic acid, modification of acyl chain length in a lipid produced by a lipid synthetic pathway comprising activity encoded by the selected shuffled lipid biosynthetic nucleic acid, location of fatty acid accumulation in a transgenic plant, algae, animal, bacteria fungus or other organism expressing the selected shuffled lipid biosynthetic nucleic acid, modulation of lipid yield of a transgenic plant, algae, animal, bacteria, fungus or other organism expressing the selected shuffled lipid biosynthetic nucleic acid, an increased ability of a molecule encoded by the selected shuffled lipid biosynthetic nucleic acid, or a cell transduced with the selected shuffled lipid biosynthetic nucleic acid, to chemically modify a lipid or lipid precursor, an increase or alteration in the range of lipid substrates for a cell transduced with the selected shuffled lipid biosynthetic nucleic acid, an increased expression level of a lipid biosynthetic polypeptide in a cell transduced with the selected shuffled lipid biosynthetic nucleic acid, a decrease in susceptibility of a lipid biosynthetic polypeptide in a cell transduced with the selected shuffled lipid biosynthetic nucleic acid to protease cleavage, a decrease in susceptibility of a lipid biosynthetic polypeptide encoded by the selected shuffled lipid biosynthetic nucleic acid in a cell to high or low pH levels, a decrease in susceptibility of a protein encoded by the selected shuffled lipid biosynthetic nucleic acid in a cell to high or low temperatures, and a decrease in toxicity to a cell by a lipid biosynthetic polypeptide

encoded by the selected shuffled lipid biosynthetic nucleic acid, as compared to one of the parental nucleic acids, when expressed in a cell.

5 The activity of the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid is selected e.g., by detecting one or more of: a change in a physical property of one or more lipid, fatty acid, wax or oil in the presence of a polypeptide or RNA encoded by the one or more lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, a protein-protein interaction in a two hybrid assay, expression of a reporter gene in a one hybrid assay, growth or survival of a recombinant cell expressing the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in an elevated temperature environment, growth or survival of a recombinant cell expressing the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a medium comprising a membrane active compound, relative bioluminescence of a recombinant cell comprising at least one gene from the Lux operon and the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid to a chloroplast, or endoplasmic reticulum, and detection of cellular localization of a product produced as a result of expression of the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a cell.

25 A variety of parental nucleic acids are suitable substrates for recombination (a nucleic acid which is to be recombined is a "substrate for recombination"), including nucleic acids which are the same as, or homologous to, a nucleic acid encoding a protein such as any of the following proteins: an Acetyl-CoA carboxylase (an ACCase), a homomeric acetyl-CoA carboxylase, a heteromeric acetyl-CoA carboxylase BC subunit, a heteromeric acetyl-CoA carboxylase, a BCCP subunit, a heteromeric acetyl-CoA carboxylase (alpha)-CT subunit, a heteromeric acetyl-CoA carboxylase (beta)-CT subunit, an acyl carrier protein (ACP) (plastidial isoform or mitochondrial isoform), a malonyl-CoA:ACP transacylase, a ketoacyl-ACP synthase

(KAS), a KAS I, a KAS II, a KAS III, a ketoacyl-ACP reductase, a 3-hydroxyacyl-ACP, an enoyl-ACP reductase, a stearoyl-ACP desaturase, an acyl-ACP thioesterase (Fat), a FatA, a FatB, a glycerol-3-phosphate acyltransferase, a 1-acyl-sn-glycerol-3-phosphate acyltransferase, a plastidial cytidine-5-diphosphate-diacylglycerol synthase, a plastidial phosphatidylglycerol-phosphate synthase, a plastidial phosphatidylglycerol-3-phosphate phosphatase, a phosphatidylglycerol desaturase (palmittate specific), a plastidial oleate desaturase (fad6), a plastidial linoleate desaturase (fad7/fad8), a plastidial phosphatidic acid phosphatase, a monogalactosyldiacylglycerol synthase, a monogalactosyldiacylglycerol desaturase (palmittate-specific), a digalactosyldiacylglycerol synthase, a sulfolipid biosynthesis protein, a long-chain acyl-CoA synthase, an ER glycerol-3-phosphate acyltransferase, an ER 1-acyl-sn-glycerol-3-phosphate acyltransferase, an ER phosphatidic acid phosphatase, a diacylglycerol cholinephosphotransferase, an ER oleate desaturase (fad2), an ER linoleate desaturase (fad3), an ER cytidine-5-diphosphate-diacylglycerol synthase, an ER phosphatidylglycerol-phosphate synthase, an ER phosphatidylglycerol-3-phosphate phosphatase, a Phosphatidylinositol synthase, a diacylglycerol kinase, a cholinephosphatase cytidyltransferase, a phosphatidylcholine transfer protein, a choline kinase, a Lipase, a phospholipase C, a phospholipase D, a phosphatidylserine decarboxylase, a phosphatidylinositol-3-kinase, a ketoacyl-CoA synthase (KCS), a (beta)-keto-acyl reductase, and a transcription factor such as CER 2 controlling lipid biosynthetic activity, a fatty acid isomerase, a fatty acid hydroxylase, a fatty acid epoxidase, a fatty acid acetyltransferase, a methyl transferase related enzyme which alters lipids, (e.g., cyclopropane fatty acid synthases, meromycolic acid synthases, cyclopropane mycolic acid synthases), a diacylglycerol acyltransferases (DGAT), an acyl-CoA reductases, a wax synthase, a Cholesterol:Acyl-CoA acyltransferases (ACAT), and/or a lecithin:Acyl-CoA Acyltransferases (LCAT).

For example, in one aspect, one or more of the parental nucleic acids are the same as, or homologous to, a nucleic acid encoding a protein which affects oil yield, such as an ACCase, an sn-2 acyltransferase, an acyltransferase other than sn-2 acyltransferase, a malonyl-CoA:ACP transacylase, an oleosin, a fatty acid binding protein, an Acyl-CoA synthase, or an acyl-ACP synthase. Similarly, at least one of the parental nucleic acids can be the same as, or homologous to, a nucleic acid encoding a protein which affects fatty acid acyl chain length or composition, such as a thioesterase

or an elongase. Again, similarly, at least one of the parental nucleic acids can be the same as, or homologous to, a nucleic acid encoding a protein which affects fatty acid saturation, such as a desaturase, a cis-trans isomerase, or a lipoxigenase (LOX). The parental nucleic acids can also be the same as, or homologous to, a nucleic acid encoding a protein which affects fatty acid branch structures, such as a reductase, or to a nucleic acid encoding a protein which affects flavor, such as a Lox protein, a desaturase, a beta-oxidation enzyme, or a hydroperoxide lyase. The parental nucleic acid can be the same as, or homologous to, a nucleic acid encoding a protein which affects polymersaturation, such as a protein in the polyketide synthase-like operon, a desaturase, or an elongase. The parental nucleic acid can be the same as, or homologous to, a nucleic acid encoding a lipase or a DNA binding protein.

A variety of recombination formats are appropriate. For example, the parental nucleic acids can be homologous or non-homologous. The parental nucleic acids can encode or not encode a lipid biosynthetic activity. The parental nucleic acids, the one or more the one or more recombinant lipid biosynthetic nucleic acid, or the selected recombinant lipid biosynthetic nucleic acid can be cloned into an expression vector.

Recombinant libraries are also an aspect of the invention. In one aspect, the plurality of parental nucleic acids are shuffled to produce a library of recombinant nucleic acids comprising one or more library member nucleic acid encoding one or more lipid biosynthetic activity. The library is optionally selected for one or more lipid biosynthetic activity such as those noted above, providing a second library. The selected library can be recombined with itself or any other nucleic acid and shuffled to produce additional recombinant selected libraries. Any of the libraries herein can be in a variety of different formats, including a phage display library, a library in a cell or cell culture such as *E. coli*, *Cynobacteria* or *Synechocystis*. The libraries can be made in a first cell type and transduced into a second cell type, e.g., the library can first be made in *E. coli* or *Cynobacteria* and then transduced into a *Synechocystis*. Another example cell type for library construction is *Pseudomonas putida*. In general, the parental nucleic acids can be shuffled in any of a plurality of cells e.g., prokaryotes or eukaryotes such as plants, yeast, bacteria, fungal cells, archae cells, or organisms.

In one format, the parental nucleic acids are shuffled in a plurality of cells and the method further includes recombining DNA from the plurality of cells that

display lipid biosynthetic activity with a library of DNA fragments, at least one of which undergoes recombination with a segment in a cellular DNA present in the cells to produce recombinant cells, or recombining DNA between the plurality of cells that display lipid biosynthetic activity to produce cells with modified lipid biosynthetic activity. In another embodiment, the method includes recombining and screening the recombinant or modified cells to produce further recombinant cells that have evolved additionally modified lipid biosynthetic activity. These steps are optionally recursively repeated until the further recombinant cells have acquired a desired lipid biosynthetic activity.

In one aspect, the method includes recombining at least one selected shuffled lipid biosynthetic nucleic acid with a further lipid biosynthetic activity nucleic acid. This further nucleic acid is the same or different from one or more of the plurality of parental nucleic acids. This recombination produces a library of recombinant lipid biosynthetic nucleic acids. The library can be screened to identify at least one further selected distinct or improved recombinant lipid biosynthetic nucleic acid that exhibits a further improvement or distinct property compared to the plurality of parental nucleic acids. Optionally, these steps are repeated until the resulting additional further distinct or improved recombinant nucleic acid shows an additionally distinct or improved lipid biosynthetic property.

In one aspect, the one or more recombinant lipid biosynthetic nucleic acid is present in one or more bacterial, yeast, plant or fungal cells and the method includes pooling multiple separate lipid biosynthetic nucleic acids. The resulting pooled lipid biosynthetic nucleic acids are screened to identify distinct or improved recombinant lipid biosynthetic nucleic acids that exhibit distinct or improved lipid biosynthetic activity compared to a non-recombinant lipid biosynthetic activity nucleic acid. The distinct or improved recombinant nucleic acid is typically cloned, transduced into a target cell or organism, or otherwise manipulated to achieve a desired effect.

One preferred recombination format is family gene shuffling. However, other shuffling protocols such as individual gene shuffling, oligonucleotide-mediated gene shuffling, in silico gene shuffling, and whole genome shuffling can be used. In addition diversity generation methods such as mutagenesis can be used to create libraries of diverse nucleic acids, separate from, or in conjunction with, shuffling methods.

One aspect of the invention is a selected shuffled lipid biosynthetic nucleic acid made by the methods herein. Similarly, a plant, bacteria or fungus transduced with the selected shuffled lipid biosynthetic nucleic acid are a feature of the invention. Because of the importance of the agricultural oils industry, plants are a preferred target for incorporation of selected nucleic acids. Preferred plants for transduction include plants in the families Gramineae, Compositae, and Leguminosae. Example preferred plants include corn, peanut, barley, millet, rice, soybean, sorghum, wheat, oats, rapeseed, oil palm, sunflower, and nut plants. The plants optionally exhibit a new lipid biosynthetic activity as compared to a wild-type non-transduced plant.

The invention also provides DNA shuffling mixtures used e.g., in the methods of the invention. For example, in one aspect, the mixture comprises at least three homologous DNAs, each of which is derived from a nucleic acid encoding a polypeptide or polypeptide fragment which encodes a lipid biosynthetic activity. As examination of the methods reveals, the at least three homologous DNAs can be present e.g., in cell culture or in vitro.

In another aspect, methods of modulating lipid biosynthetic activity in a cell are provided. In these methods, whole genome shuffling of a plurality of genomic nucleic acids in the cell is performed and one or more lipid biosynthetic activity is selected. The genomic nucleic acids can be from a species or strain the same as or different from the cell, which may be of, e.g., eukaryotic or prokaryotic origin. Any of the lipid biosynthetic activities noted above can be selected.

In another aspect, the invention provides methods of obtaining a recombinant lipid biosynthetic nucleic acid which can confer modified lipid production to a plant in which the recombinant lipid biosynthetic nucleic acid is present. In method, a plurality of forms of a selected lipid synthetic nucleic acid are recombined. These forms include segments derived from one or more parental nucleic acid which encode a lipid biosynthetic activity, or which can be shuffled to confer a lipid biosynthetic activity. The plurality of forms of the selected nucleic acid differ from each other in at least one nucleotide. This recombination produces a library of recombinant lipid biosynthetic nucleic acids. The library is screened to identify at least one recombinant lipid biosynthetic nucleic acid that exhibits distinct or improved lipid biosynthetic activity as compared to the parental nucleic acid. Typically, one or more parental nucleic acid encodes a lipid biosynthetic enzyme, although the parental nucleic

acids can encode other factors which affect lipid synthesis, or which can be shuffled to affect lipid synthesis. Any of the biosynthetic coding sequences and selection procedures noted above are applicable to this method.

In any of the methods, the libraries produced can be screened by any of a variety of methods, such as selection by growing cells comprising the library in or on a medium comprising a cell membrane disruptive agent. The libraries are optionally selected for one or more additional lipid biosynthetic activity.

The step of recombining cells of a library is optionally performed in a plurality of cells. This recombination can include recombining DNA from the plurality of cells that display a selected lipid biosynthetic phenotype with a second library of DNA fragments, at least one of which undergoes recombination with a segment in a nucleic acid present in the cells to produce recombined modified lipid synthetic cells, or recombining DNA between the plurality of cells that display a selected lipid biosynthetic phenotype to produce modified lipid synthetic cells. The library can be recombined and screened to produce further recombined cells that have evolved additionally distinct or improved lipid synthetic activity. These steps can be iteratively repeated.

#### BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 is a diagram of fatty acid synthesis in plants.

Fig. 2 is a diagram of elongation of the acyl group in the fatty acid synthase cycle.

Fig. 3A-3B is a schematic of a hybrid protein assay.

Fig. 4 is a schematic of a LUX assay.

#### DEFINITIONS

Unless clearly indicated to the contrary, the following definitions supplement definitions of terms known in the art.

A "recombinant" nucleic acid is a nucleic acid produced by recombination between two or more nucleic acids, or any nucleic acid made by an *in vitro* or artificial process. The term "recombinant" when used with reference to a cell indicates that the cell comprises (and optionally replicates) a heterologous nucleic acid, or expresses a peptide or protein encoded by a heterologous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) parental

form of the cell. Recombinant cells can also contain genes found in the native form of the cell where the genes are modified and re-introduced into the cell by artificial means. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been artificially modified without removing the nucleic acid from the cell, such modifications include those obtained by gene replacement, site-specific mutation, and related techniques.

A "lipid" is a molecule which is fat soluble (e.g., soluble in a non-polar solvent), such as a fatty acid, glyceride, glyceryl ether, phospholipid, sphingolipid, alcohol, wax, oil, terpene, steroid, fat soluble vitamin, or the like.

A "recombinant lipid biosynthetic nucleic acid" is a recombinant nucleic acid encoding a protein, RNA (or other active nucleic acid) which produces one or more lipid, *in vitro* or *in vivo* or which interacts with one or more additional RNAs (or other active nucleic acids, such as DNAs in the case where the recombinant lipid biosynthetic nucleic acid encodes a transcription factor) or proteins *in vitro* or *in vivo* to produce one or more lipid. Similarly, a lipid biosynthetic activity is any activity which produces or controls production of one or more lipid (including both yield and lipid type).

A "plurality of forms" of a selected nucleic acid refers to a plurality of homologs of the nucleic acid. The homologs can be from naturally occurring homologs (e.g., two or more homologous genes) or by artificial synthesis of one or more nucleic acid(s) having related sequences (e.g., as occurs during oligonucleotide-mediated family gene shuffling), or by modification of one or more nucleic acid to produce related nucleic acids. Nucleic acids are homologous when they are derived, naturally or artificially, from a common ancestor sequence. During natural evolution, this occurs when two or more descendant sequences diverge from a parent sequence over time, i.e., due to mutation and natural selection. Under artificial conditions, divergence occurs, e.g., in one of two ways. First, a given sequence can be artificially recombined with another sequence, as occurs, e.g., during typical cloning, to produce a descendant nucleic acid. Alternatively, a nucleic acid can be synthesized *de novo*, by synthesizing a nucleic acid which varies in sequence from a given parental nucleic acid sequence.

When there is no explicit knowledge about the ancestry of two nucleic acids, homology is typically inferred by sequence comparison between two sequences. Where two nucleic acid sequences show sequence similarity it is inferred that the two nucleic acids share a common ancestor. The precise level of sequence similarity

required to establish homology varies in the art depending on a variety of factors. For purposes of this disclosure, two sequences are considered homologous, e.g., where they share sufficient sequence identity to allow recombination to occur between two nucleic acid molecules. Typically, nucleic acids require regions of close similarity spaced roughly the same distance apart to permit recombination to occur. Typically regions of at least about 60% sequence identity or higher are optimal for recombination.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to two or more sequences or subsequences that have at least about 60%, preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Such "substantially identical" sequences are typically considered to be homologous.

Preferably, the "substantial identity" exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad.*

*Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel *et al.*, *infra*). One example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length *W* in the query sequence, which either match or satisfy some positive-valued threshold score *T* when aligned with a word of the same length in a database sequence. *T* is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*).

These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters *M* (reward score for a pair of matching residues; always > 0) and *N* (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity *X* from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters *W*, *T*, and *X* determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (*W*) of 11, an expectation (*E*) of 10, a cutoff of 100, *M*=5, *N*=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (*W*) of 3, an expectation (*E*) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (*P*(*N*)), which provides an indication of the probability by which a match between two

nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Another indication that two nucleic acid sequences are substantially identical/homologous is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to," refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions, including when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target polynucleotide sequence. "Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes* part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays," Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5° C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. Typically, under "stringent conditions" a probe will hybridize to its target subsequence, but not to unrelated sequences.

The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the  $T_m$  for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42° C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at 72° C for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC

wash at 65° C for 15 minutes (*see*, Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 45° C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6x SSC at 40° C for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30° C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

A "subsequence" refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (e.g., polypeptide) respectively.

The term "gene" is used broadly to refer to any segment of nucleic acid (typically DNA, but optionally RNA) associated with expression of a given RNA or protein. Thus, genes include sequences encoding expressed RNAs (which can include polypeptide coding sequences) and, often, the regulatory sequences required for their expression. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

The term "isolated", when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state.

The term "nucleic acid" refers to deoxynucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. The term nucleic acid is generic to the terms "gene", "DNA", "cDNA", "oligonucleotide", "RNA", "mRNA", and the like.

"Nucleic acid derived from a gene" refers to a nucleic acid for whose synthesis the gene, or a subsequence thereof, has ultimately served as a template. Thus, an mRNA, a cDNA reverse transcribed from an mRNA, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, etc., are all derived from the gene and detection of such derived products is indicative of the presence and/or abundance of the original gene and/or gene transcript in a sample.

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it increases the transcription of the coding sequence.

A "recombinant expression cassette" or simply an "expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with nucleic acid elements that are capable of effecting expression of a structural gene in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes a nucleic acid to be transcribed (e.g., a nucleic acid encoding a desired polypeptide), and a promoter. Additional factors necessary or helpful in effecting expression may also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell. Transcription termination signals, enhancers, and other nucleic acid sequences that influence gene expression, can also be included in an expression cassette.

#### DETAILED DISCUSSION OF THE INVENTION

The present invention provides ways of improving or modulating lipid production in cells and whole organisms such as oil crop plants. Examples of fatty acid modifying phenotypes include, but are not limited to, an increase or decrease in level of saturation of the fatty acid, the length of the carbon backbone, the position of double or triple bonds, the linkage forms of the fatty acyl groups, the location of fatty acid accumulation, and the quantity (yield) of total triglycerides and other forms of lipid in the organism. DNA sequences that can be shuffled according to the present invention include, but are not limited to, the exons and introns of a gene, promoter sequences controlling expression of genes, transcription factors regulating gene expression,

operons (an "operon" is typically a group of contiguous structural genes which are transcribed as a single transcription unit from a common promoter and can be thereby subject to coordinated regulation) involved in a fatty acid production pathway and the like.

In addition, a complete fatty acid synthetic pathway from plants or other organisms, including bacteria, fungi, algae and animals, can be linked together through cloning to result in a polypeptide (Halpin, C. et al. 1999. Self-processing 2A-polypeptides - a system for co-ordinate expression of multiple proteins in transgenic plants (*Plant J* 17(2): 453-459). Such polypeptide genes can be subject to shuffling in the individual gene segment or the entire gene. These polypeptide genes or T-DNAs containing multiple fatty acid-related genes (under control of one or more promoters) can be introduced into plant chromosomes or plasmid genomes by conventional plant gene transformation methods.

Furthermore, genes involved in fatty acid synthesis in animal, bacteria or fungi, many of which are incompatible with the plant fatty acid synthetic pathway due to different substrate preferences, are modified to utilize substrates abundant in the plant pathway. As a result of DNA shuffling, the encoded product can have a modified property, such as increase or decrease of the substrate specificity when compared to the wild-type protein. In some applications, the enzyme is altered to recognize a substrate other than its natural one, resulting in a novel phenotype in a desired host.

For example, in plant seeds, fatty acid synthesis mainly occurs in the plastids and most enzymes involved utilize acyl-ACP, but not free fatty acids, as substrates (see Fig. 2). In one embodiment, a bacterial enzyme utilizing free fatty acid to produce a desired product is shuffled to recognize acyl-ACP and thus can be transformed into plants to produce a novel phenotype. Furthermore, the production of certain fatty acids are controlled by a gene operon, in which the acyl group is channeled from one gene product to the next. It is often not clear which gene product is rate-limiting. The complete operon can be shuffled as a whole, resulting in a much more efficient system for fatty acid production of one or more fatty acids from any of a variety of substrates.

## DNA SHUFFLING FOR LIPID MODIFICATIONS

In general, biotechnological approaches to conferring desirable lipid production to crops involves either: (a) altering the gene that codes for a target site in order to confer desirable properties, or (b) engineering a gene into crops that codes for an enzyme with a desirable property. Traditionally, such enzymes are discovered either by extensive screening of organisms or by mutagenesis followed by rigorous selection. In spite of this rigorous scheme, selected enzymes may not have the ideal properties to confer crop selectivity or to function effectively in transgenic crops, and the process is, at best, labor intensive.

The present invention overcomes these difficulties by applying DNA shuffling and other diversity generation/ selection techniques to gene-families that code for lipid synthesis or metabolizing genes, including those listed herein. Such genes are optimized by DNA shuffling in order to enhance the rate of metabolism or synthesis of specific lipids or lipid substrates, optionally without altering other parameters, such as affinity for natural substrates, effectors, etc., or, alternately, optionally including altering these additional parameters. A number of specific applications are given herein by way of example.

## FATTY ACID ANALYSIS

Fatty acid compositions can be analyzed using several established protocols. For example, plant seed fatty acid composition may be determined by the acid methanolysis method described by Browse et al. (*Anal. Biochem.* 1986, 152: 141-145). In other cases where a large number of samples are involved, to identify novel fatty acids produced the shuffled genes or pathways is performed using high throughput assays. Fatty acids, phospholipids and triglycerides are detected, e.g., using ESI (electrospray ionization) or APCI (atmospheric pressure chemical ionization) mass spectrometry (Karlsson, A.A. et al., *J. Mass Spectrom.*, 1998, Hoischen, C. et al. I. *Bacteriol.*, 1997). In comparison to the traditional GC/MS detection, these methods allow for direct screening of the analyte molecule without prior derivatization. Gas chromatography based approaches are also applicable in the present invention for screening fatty acids.

A high throughput method for detecting analyte molecules from a complex biological matrix by electrospray tandem mass spectrometry is taught in

"HIGH THROUGHPUT MASS SPECTROMETRY" by Sun Ai Railand, USSN 60/119,766, filed 02/11/1999, which utilizes off-line parallel sample purifications and fast flow-injection analysis, typically reducing the time of analysis to 30 to 40 seconds per sample. All steps starting from cell picking, cell growth, sample preparation and analysis are automated and can be carried out overnight by various robotic workstations.

Tandem mass spectrometry allows for high selectivity and sensitivity and for simultaneous detection of multiple analytes. The analysis by mass spectrometry allows for identification based on mass over charge. Tandem mass spectrometry can potentially distinguish regioisomers as well. For further analysis of the structure, one can employ UV (for conjugation of double bonds) or capillary electrophoresis separation system with UV or fluorescent detection (Akaeoka, K. et al. *Enantiomer.* 1998).

A number of additional high-throughput diversity generation (e.g., by recombination or mutation) and integrated screening systems which are applicable to the present invention are described in Bass et al. "Integrated Systems and Methods for Diversity Generation and Screening" filed January 11, 2000, USSN 60/175,551.

## FATTY ACID BIOSYNTHETIC PATHWAYS AND ENZYMES INVOLVED

A variety of the enzymes and corresponding genes for fatty acid synthesis and metabolism are known. For example, the pathway set forth in Fig. 1 provides an example fatty acid synthesis cycle in plants.

The details of the fatty acid synthase cycle is shown in Fig. 2. The enzymes involved are numbered and listed below. Figure 2 shows elongation of acyl groups in the fatty acid synthase cycle.

Enzymes in the pathway include: 1.- Acetyl-CoA carboxylase (ACCase); 1a.- Homomeric acetyl-CoA carboxylase (EC 6.4.1.2); 1b.- Heteromeric acetyl-CoA carboxylase BC subunit (EC 6.4.1.2); 1c.- Heteromeric acetyl-CoA carboxylase BCCP subunit (EC 6.4.1.2); 1d.- Heteromeric acetyl-CoA carboxylase (alpha)-CT subunit (EC 6.4.1.2); 1e.- Heteromeric acetyl-CoA carboxylase (beta)-CT subunit (EC 6.4.1.2); 2.- Acyl carrier proteins (ACP) plastidial isoforms mitochondrial isoforms; 3.- Malonyl-CoA:ACP transacylase (EC 2.3.1.39); 4.- Ketosacyl-ACP synthase (KAS); 4a.- KAS I (EC 2.3.1.41); 4b.- KAS II (EC 2.3.1.41); 4c.- KAS III (EC

2.3.1.41); 5- Ketosyl-ACP reductase (EC 1.1.1.100); 6-3-hydroxyacyl-ACP  
 dehydratase (EC 4.2.1.17); 7- Enoyl-ACP reductase (EC 1.3.1.44); 8- Stearoyl-ACP  
 desaturase (EC 1.14.99.6); 9- Acyl-ACP thioesterase (Fad) (EC 3.1.2.14); 9a- Fatty;  
 9b- FAD; 10- glycerol-3-phosphate acyltransferase (EC 2.3.1.5); 11-1-acyl-sn-  
 glycerol-3-phosphate acyltransferase (EC 2.3.1.51); 12- Plasmalogen cytidine-5'-  
 diphosphate-diacylglycerol synthase (EC 2.7.7.41); 13- Plasmalogen phosphatidylglycero-  
 phosphate synthase; 14- Plasmalogen phosphatidylglycerol-3-phosphate phosphatase; 15-  
 Phosphatidylglycerol desaturase (palmitate specific) (EC 1.14.99.-); 16- Plasmalogen  
 oleate desaturase (fad6) (EC 1.14.99.-); 17- Plasmalogen linoleate desaturase (fad7/fad8)  
 (EC 1.14.99.-); 18- Plasmalogen phosphatidic acid phosphatase (EC 3.1.3.4); 19-  
 Monogalactosyl-diacyl-glycerol synthase (EC 2.4.1.46); 20- Monogalactosyl-diacyl-  
 glycerol desaturase (palmitate specific) (EC 1.14.99.-); 21- Digalactosyl-diacyl-glycerol  
 synthase (EC 2.4.1.184); 22- Sulfolipid biosynthesis protein; 23- Long-chain acyl-CoA  
 synthase (EC 6.2.1.3); 24- ER glycerol-3-phosphate acyltransferase; 25- ER 1-acyl-  
 sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51); 26- ER phosphatidic acid  
 phosphatase; 27- Diacylglycerol cholinephosphotransferase (EC 2.7.8.2); 28- ER  
 oleate desaturase (fad2) (EC 1.14.99.-); 29- ER linoleate desaturase (fad3) (EC  
 1.14.99.-); 30- ER cytidine-5'-diphosphate-diacylglycerol synthase (EC 2.7.7.41); 31-  
 ER phosphatidylglycero-phosphate synthase; 32- ER phosphatidylglycerol-3-phosphate  
 phosphatase; 33- Phosphatidylinositol synthase (EC 2.7.8.11).

The following enzymes/reactions are not shown in Figures 1 or 2, but  
 can also be evolved as set forth herein: Related to linoleoyl desaturase Diacylglycerol  
 kinase (EC 2.7.1.107); Cholinephosphate cytidylyltransferase (EC 2.7.7.15); Similar to  
 acyl-CoA desaturase (EC 1.14.99.-); Phosphatidylcholine Transfer Protein; Choline  
 kinase (EC 2.7.1.32); Lipases; Phospholipase C (EC 3.1.4.11); Phospholipase D (EC  
 3.1.4.4); Phosphatidylserine decarboxylase (EC 4.1.1.65); Phosphatidylinositol-3-  
 kinase (EC 2.7.1.137); Ketosyl-CoA synthase (KCS); (beta)-keto-acyl reductase  
 (involved in wax biosynthesis); Putative transcription factor CER2 involved in wax  
 biosynthesis; Fatty acid isomerase; Fatty acid hydroxylase; Fatty acid epoxidase; Fatty  
 acid acetylase; Methyl transferase related enzyme which alters lipids (e.g.,  
 cyclopropane fatty acid synthases, meromycolic acid synthases, cyclopropane mycolic  
 acid synthases); Diacylglycerol acyltransferases (DGAT); acyl Co-A reductases; wax

synthases; Cholesterol:Acyl-CoA acyltransferases (ACAT); and/or a lecithin:Acyl-CoA  
 Acyltransferases (LCAT).

#### EXAMPLE TARGET ENZYMES/SOURCES

Many of the genes and biochemical pathways for lipid biosynthetic  
 genes and enzymes listed above, as well as many others, are known, and can be found  
 in various sequence repositories such as GenBank. Examples of some target enzymes  
 and their gene sources include the following:

##### (1) Oil yield:

ACCase (Reverdatto S, 1999, *Plant Physiol.* 119:961-978); sn-2  
 acyltransferase (Knutzon DS et al. 1995 *Plant Physiol.* 109:999-1006); sn-2

acyltransferases (for example, Lassner MW et al. 1995, *Plant Physiol.* 109:1389-1394);  
 malonyl-CoA:ACP transacylase (Verwoert II, Et al. 1992, *J. Bacteriol.* 174:2851-2857;  
 Summers RG et al. 1995, *Biochem.* 34: 9389-9402); oleosins (Parmenter DL et al.  
 1995, *Plant Mol Biol.* 29:1167-1180); fatty acid binding proteins (Castagnaro A and  
 Garcia-Olmedo F, 1994 *FEBS Lett.* 349: 117-119); Acyl-CoA synthase (Choi et al.  
 1999, *J Biol Behav* 274:4671-4682); Acyl-ACP synthase (Juckowski S et al. 1994 *J*  
*Biol. Chem* 269:2921-2928);

##### (2) Chain length:

Thioesterases (Voelker TA et al. 1992, *Science.* 257: 72-74; Ferni SR  
 and Meighen EA, 1991, *J Biol. Chem.* 266: 12852-12857); elongases (KAS) (Lassner  
 MW, et al. 1996, *Plant Cell* 8:281-292);

##### (3) Saturation:

Desaturases (Knutzon DS et al. 1992, *Proc. Nat Acad Sci USA.* 89:  
 2624-2628); cis-trans isomerase (Loftfield B and Keweleob H, 1996, *Lipids.* 31:811-815);  
 lipoygenase (Lox) (Kausch KD and Handa AK, 1997, *Plant Physiol.* 113:1041-1050);

##### (4) Branch:

Reductases (Wallace KK et al. 1995, *Eur J Biochem.* 233:954-962;  
 Duran E et al. 1993 *J. Biol. Chem.* 268:22391-22396);

##### (5) Flavor:

Lox; desaturases; beta-oxidation enzymes (Bojorguez G et al. 1995 *Plant*  
*Mol Biol* 28: 811-820); Hydroperoxide lyase (Bate NJ et al. 1998 *Plant Physiol.*  
 117:11393-1400); Matsui K et al. 1996, *FEBS Lett* 394: 21-24);

6) PUFA (Polyunsaturated fatty acids): PKs-like opetron, Desaturases (Reddy AS and Thomas TL. 1996. Nat Biotechnol. 14: 639-642); Elongases;

5 (7) Lipases  
(Brick DJ. Et al. 1995 FEBS Lett 377:475-480);

(8) DNA-binding proteins  
(Oskoui B and Saba JD 1999. Mol Gen Genet. 261:346-353).

## GENERAL DNA SHUFFLING FORMATS

The invention provides significant advantages over previously used methods for optimization of lipid biosynthesis and metabolic genes. For example, DNA shuffling can result in optimization of a desirable property even in the absence of a detailed understanding of the mechanism by which the particular property is mediated (this is especially the case for whole-genome shuffling formats where even the targets for shuffling can be completely unknown). In addition, entirely new properties can be obtained upon shuffling of DNAs, i.e., shuffled DNAs can encode polypeptides or RNAs with properties entirely absent in the parental DNAs which are shuffled.

Sequence recombination can be achieved in many different formats and permutations of formats, as described in further detail below. These formats share some common principles.

20 The substrates for modification, or “forced evolution,” vary in different applications, as does the property sought to be acquired or improved. Examples of candidate substrates for acquisition of a property or improvement in a property include genes (typically the regulatory elements directing expression of a coding sequence in a cell or *in vitro* transcription reaction) that encode proteins which have enzymatic or other activities useful forming linkages in lipid molecules, in breaking down lipid molecules, in sequestering lipid molecules and the like.

The methods typically use at least two variant forms of a starting substrate. The variant forms of candidate substrates can show substantial sequence or secondary structural similarity with each other, but they should also differ in at least one and preferably at least two positions. The initial diversity between forms can be the result of natural variation, *e.g.*, the different variant forms (homologs) are obtained from different individuals or strains of an organism (including geographic variants) or constitute related sequences from the same organism (*e.g.*, allelic variations), or

constitute homologs from different organisms (interspecific variants). Shuffling of such natural variants is one form of "family gene shuffling."

Alternatively, initial diversity can be induced, e.g., the variant forms can be generated by error-prone transcription, such as an error-prone PCR or use of a polymerase which lacks proof-reading activity (see, Liao (1990) *Gene* 88: 107-111), of the first variant form, or, by replication of the first form in a mutator strain (mutator host cells are discussed in further detail below, and are generally well known). The initial diversity between substrates is greatly augmented in subsequent steps of recombination for library generation. A mutator strain can include any mutants in any organism impaired in the functions of mismatch repair. These include mutant gene products of *mutS*, *mutT*, *mutH*, *mutL*, *ovdD*, *dcm*, *var*, *umuC*, *umuD*, *soxB*, *recJ*, etc. The impairment is achieved by genetic mutation, allelic replacement, selective inhibition by an added reagent such as a small compound or an expressed antisense RNA, or other techniques. Impairment can be of the genes noted, or of homologous genes in any organism.

The properties or characteristics that can be acquired or improved vary widely; and, of course, depend on the choice of substrate. For example, for lipid synthetic genes, properties that one can improve include, but are not limited to a change in a physical property of one or more lipid, fatty acid, wax or oil in the presence of a polypeptide or RNA encoded by the one or more lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, a protein-protein interaction in a two hybrid assay, expression of a reporter gene in a one hybrid assay, growth or survival of a recombinant cell expressing the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in an elevated temperature environment, growth or survival of a recombinant cell expressing the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a medium comprising a membrane active compound, relative bioluminescence of a recombinant cell comprising at least one gene from the Lux operon and the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected

shuffled lipid biosynthetic nucleic acid to a chloroplast, or endoplasmic reticulum, and detection of cellular localization of a product produced as a result of expression of the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a cell.

At least two variant forms of a nucleic acid which can confer lipid synthetic or metabolic activity are recombined to produce a library of recombinant nucleic acids. The library is then screened to identify at least one recombinant nucleic acid that is optimized for the particular property or properties of interest.

Often, improvements are achieved after one round of recombination and selection. However, recursive sequence recombination can be employed to achieve still further improvements in a desired property, or to bring about new (or "distinct") properties. Recursive sequence recombination entails successive cycles of recombination to generate molecular diversity. That is, one creates a family of nucleic acid molecules showing some sequence identity to each other but differing in the presence of mutations. In any given cycle, recombination can occur *in vivo* or *in vitro*, intracellularly or extracellularly. Furthermore, diversity resulting from recombination can be augmented in any cycle by applying prior methods of mutagenesis (e.g., error-prone PCR or cassette mutagenesis) to either the substrates or products for recombination.

A recombination cycle is usually followed by at least one cycle of screening or selection for molecules having a desired property or characteristic. If a recombination cycle is performed *in vitro*, the products of recombination, i.e., recombinant segments, are sometimes introduced into cells before the screening step. Recombinant segments can also be linked to an appropriate vector or other regulatory sequences before screening. Alternatively, products of recombination generated *in vitro* are sometimes packaged in viruses (e.g., bacteriophage) before screening. If recombination is performed *in vivo*, recombination products can sometimes be screened in the cells in which recombination occurred. In other applications, recombinant segments are extracted from the cells, and optionally packaged as viruses, before screening.

The nature of screening or selection depends on what property or characteristic is to be acquired or the property or characteristic for which improvement is sought, and many examples are discussed below. It is not usually necessary to

23

understand the molecular basis by which particular products of recombination (recombinant segments) have acquired new or improved properties or characteristics relative to the starting substrates. For example, a lipid biosynthetic gene can have many component sequences each having a different intended role (e.g., coding sequence, regulatory sequences, targeting sequences, stability-conferring sequences, and sequences affecting integration). Each of these component sequences can be varied and recombined simultaneously. Screening/selection can then be performed, for example, for recombinant segments that have increased ability to confer lipid synthetic traits to a plant without the need to attribute such improvement to any of the individual component sequences of the vector.

Depending on the particular screening protocol used for a desired property, initial round(s) of screening can sometimes be performed using bacterial cells due to high transfection efficiencies and ease of culture. Later rounds, and other types of screening which are not amenable to screening in bacterial cells, can be performed, e.g., in plant cells to optimize recombinant segments for use in an environment close to that of their intended use. Final rounds of screening can be performed in the precise cell type of intended use (e.g., a cell which is present in a plant), or even in whole plants (e.g., crop tests in the field) or other organisms.

In some methods, use of a recombinant gene can itself be used as a round of screening. That is, recombinant genes that are successfully taken up and/or expressed by the intended target cells are recovered from those target cells and used to confer traits upon other plants. The recombinant genes that are recovered from the first target cells are enriched for genes that have evolved, i.e., have been modified by recursive sequence recombination, toward improved or new properties or characteristics for specific uptake and integration of the gene, desired lipid levels, stability, and the like.

The screening or selection step identifies a subpopulation of recombinant segments that have evolved toward acquisition of a new or improved desired property or properties useful in conferring lipid synthetic activity upon plants. Depending on the screen, the recombinant segments can be identified as components of cells, components of viruses or in free form. More than one round of screening or selection can be performed after each round of recombination.

24

If further improvement in a property is desired, at least one and usually a collection of recombinant segments surviving a first round of screening/selection are subject to a further round of recombination. These recombinant segments can be recombined with each other or with exogenous segments representing the original substrates or further variants thereof. Again, recombination can proceed *in vitro* or *in vivo*. If the previous screening step identifies desired recombinant segments as components of cells, the components can be subjected to further recombination *in vivo*, or can be subjected to further recombination *in vitro*, or can be isolated before performing a round of *in vitro* recombination. Conversely, if the previous screening step identifies desired recombinant segments in naked form or as components of viruses, these segments can be introduced into cells to perform a round of *in vivo* recombination. The second round of recombination, irrespective how performed, generates further recombinant segments which encompass additional diversity than is present in recombinant segments resulting from previous rounds.

The second round of recombination can be followed by a further round of screening/selection according to the principles discussed above for the first round. The stringency of screening/selection can be increased between rounds. Also, the nature of the screen and the property being screened for can vary between rounds if improvement in more than one property is desired or if acquiring more than one new property is desired. Additional rounds of recombination and screening can then be performed until the recombinant segments have sufficiently evolved to acquire the desired new or improved property or function.

The practice of this invention involves the construction of recombinant nucleic acids and the expression of genes in transfected host cells. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids such as expression vectors are well-known to persons of skill. General texts which describe molecular biological techniques useful herein, including mutagenesis, include Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, CA (Berger), Sambrook et al., *Molecular Cloning - A Laboratory Manual* (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and *Current Protocols in Molecular Biology*, F.M. Ausubel et al., eds., Current Protocols, a joint venture between

Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1998) ("Ausubel"). Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q $\beta$ -replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA) are found in Berger, Sambrook, and Ausubel, as well as Mullis et al., (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3, 81-94; (Kwoh et al. (1989) *Proc. Natl. Acad. Sci. USA* 86, 1173; Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell et al. (1989) *J. Clin. Chem* 35, 1826; Landegren et al., (1988) *Science* 241, 1077-1080; Van Brunt (1990) *Biotechnology* 8, 291-294; Wu and Wallace, (1989) *Gene* 4, 560; Bartinger et al. (1990) *Gene* 89, 117, and Sooknahan and Malek (1995) *Biotechnology* 13: 563-564.

Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) *Nature* 369: 684-685 and the references therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, Ausubel, Sambrook and Berger, *all supra*.

Oligonucleotides for use as probes, e.g., in *in vitro* amplification methods, for use as gene probes, or as shuffling targets (e.g., synthetic genes or gene segments) are typically synthesized chemically according to the solid phase phosphoramidite trister method described by Beaucage and Caruthers (1981), *Terrahedron Letts*, 22(20):1859-1862, e.g., using an automated synthesizer, as described in Niedham-Vandevanter et al. (1984) *Nucleic Acids Res.* 12:6159-6168. Oligonucleotides can also be custom made and ordered from a variety of commercial sources known to persons of skill.

#### AUTOMATION FOR STRAIN IMPROVEMENT

One aid to strain improvement is having an assay that can be dependably used to identify a few mutants out of thousands that have potentially subtle increases in product yield. The limiting factor in many assay formats is the uniformity of library

cell (or viral) growth. This variation is the source of baseline variability in subsequent assays. Inoculum size and culture environment (temperature/humidity) are sources of cell growth variation. Automation of all aspects of establishing initial cultures and state-of-the-art temperature and humidity controlled incubators are useful in reducing variability.

In one aspect, library members, e.g., cells, viral plaques, spores or the like, are separated on solid media to produce individual colonies (or plaques). Using an automated colony picker (e.g., the Q-bot, Genetix, U.K.), colonies are identified, picked, and 10,000 different mutants inoculated into 96 well microtiter dishes containing two 3 mm glass balls/well. The Q-bot does not pick an entire colony but rather inserts a pin through the center of the colony and exits with a small sampling of cells, (or mycelia) and spores (or viruses in plaque applications). The time the pin is in the colony, the number of dips to inoculate the culture medium, and the time the pin is in that medium each effect inoculum size, and each can be controlled and optimized.

15 The uniform process of the Q-bot decreases human handling error and increases the rate of establishing cultures (roughly 10,000/4 hours). These cultures are then shaken in a temperature and humidity controlled incubator. The glass balls in the microtiter plates act to promote uniform aeration of cells and the dispersal of mycelial fragments similar to the blades of a fermenter.

#### 20 (a.) Prescreen

The ability to detect a subtle increase in the performance of a shuffled library member over that of a parent strain relies on the sensitivity of the assay. The chance of finding the organisms having an improvement is increased by the number of individual mutants that can be screened by the assay. To increase the chances of identifying a pool of sufficient size, a prescreen that increases the number of mutants processed by 10-fold can be used. The goal of the primary screen is to quickly identify mutants having equal or better product titers than the parent strain(s) and to move only these mutants forward to liquid cell culture for subsequent analysis.

25 An especially preferred high throughput method for detecting analyte molecules from a complex biological matrix is by electrospray tandem mass spectrometry as taught in "HIGH THROUGHPUT MASS SPECTROMETRY" by Sam Ai Railhard, USSN 60/19,766, filed 02/11/1999. In the '766 application, methods which utilize off-line parallel sample purification and fast flow-injection analysis,

typically reducing the time of analysis to 30 to 40 seconds per sample. All steps starting from cell picking, cell growth, sample preparation and analysis are automated and can be carried out overnight by various robotic workstations.

#### 5 FORMATS FOR LIBRARY DIVERSITY GENERATION INCLUDING SEQUENCE RECOMBINATION

As described herein, the methods of the invention optionally entail performing recombination ("shuffling") or other sequence diversity generation protocols (e.g., mutation) and screening or selection to "evolve" individual genes, whole plasmids or viruses, multigene clusters, or even whole genomes (*see also*, Stemmer (1995) *BioTechnology* 13:549-553, for an introduction to shuffling).

##### 10 Iterative cycles of diversity generation/ recombination and

screening/selection can be performed to further evolve the nucleic acids of interest. Such techniques do not require the extensive analysis and computation required by conventional methods for polypeptide engineering. Shuffling allows the recombination of large numbers of mutations in a minimum number of selection cycles, in contrast to natural pairwise recombination events (e.g., as occur during sexual replication). Thus, the sequence recombination techniques described herein provide particular advantages in that they provide recombination between mutations in any or all of these, thereby providing a very fast way of exploring the manner in which different combinations of mutations can affect a desired result. In some instances, however, structural and/or functional information is available which, although not required for sequence recombination, provides opportunities for modification of the technique.

20 Generally, shuffling procedures can be used in conjunction with other diversity generation protocols for generating libraries of diverse A variety of diversity generating protocols, including nucleic acid shuffling protocols, including nucleic acid family shuffling protocols and other specific desirable recombination formats are available and described in the art. The following publications describe a variety of recursive recombination procedures and/or methods which can be incorporated into such procedures, as well as other diversity generating protocols: Stemmer, et al., (1999) "Molecular breeding of viruses for targeting and other clinical properties." *Tumor Targeting* 4:1-4; Nasset et al. (1999) "DNA Shuffling of subgenomic sequences of subtilisin." *Nature Biotechnology* 17:893-896; Chang et al. (1999) "Evolution of a cytokine using DNA family shuffling." *Nature Biotechnology* 17:793-797; Minshall and

- Stemmer (1999) "Protein evolution by molecular breeding" Current Opinion in Chemical Biology 3:284-290; Christians et al. (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" Nature Biotechnology 17:229-264; Cramer et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature 391:288-291; Cramer et al. (1997) "Molecular evolution of an ascorbate detoxification pathway by DNA shuffling" Nature Biotechnology 15:436-438; Zhang et al. (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" Proceedings of the National Academy of Sciences, U.S.A. 94:4504-4509; Patten et al. (1997)
- 10 "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" Current Opinion in Biotechnology 8:724-733; Cramer et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" Nature Medicine 2:100-103; Cramer et al. (1996) "Improved green fluorescent protein by molecular evolution using DNA shuffling" Nature Biotechnology 14:315-319; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor headpiece dimer" Journal of Molecular Biology 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: The Encyclopedia of Molecular Biology. VCH Publishers, New York. pp.447-457; Cramer and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes"
- 20 Biotecniques 18:194-195; Stemmer et al. (1995) "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxynucleotides" Gene. 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" Science 270: 1510; Stemmer (1995) "Searching Sequence Space" BioTechnology 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" Nature 370:389-391; and Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." Proceedings of the National Academy of Sciences, U.S.A. 91:10747-10751.

Additional details regarding DNA shuffling and other diversity generating methods and compositions are found in U.S. Patents by the inventors and their co-workers, including: United States Patent 5,605,793 to Stemmer (February 25, 1997), "METHODS FOR IN VITRO RECOMBINATION;" United States Patent 5,811,238 to Stemmer et al. (September 22, 1998) "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE

SELECTION AND RECOMBINATION;" United States Patent 5,830,721 to Stemmer et al. (November 3, 1998), "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY;" United States Patent 5,834,252 to Stemmer, et al. (November 10, 1998) "END-COMPLEMENTARY POLYMERASE REACTION;" and United States Patent 5,837,458 to Minsahl, et al. (November 17, 1998), "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING."

In addition, details and formats for DNA shuffling and other diversity generating protocols are found in a variety of PCT and foreign patent application publications, including: Stemmer and Cramer, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" WO 95/22625; Stemmer and Lipschultz "END-COMPLEMENTARY POLYMERASE CHAIN REACTION" WO 96/33207; Stemmer and Cramer "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND RECOMBINATION" WO 97/0078; Minsahl and Stemmer, "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING" WO 97/35966; Punnonen et al. "TARGETING OF GENETIC VACCINE VECTORS" WO 99/41402; Punnonen et al. "ANTIGEN LIBRARY IMMUNIZATION" WO 99/41383; Punnonen et al. "GENETIC VACCINE VECTOR ENGINEERING" WO 99/41369; Punnonen et al. "OPTIMIZATION OF IMMUNOMODULATORY PROPERTIES OF GENETIC VACCINES" WO 99/41368; Stemmer and Cramer, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" EP 0934999; Stemmer "EVOLVING CELLULAR DNA UPTAKE BY RECURSIVE SEQUENCE RECOMBINATION" EP 0932670; Stemmer et al., "MODIFICATION OF VIRUS TROPISM AND HOST RANGE BY VIRAL GENOME SHUFFLING" WO 99/23107; Api et al., "HUMAN PAPILLOMA VIRUS VECTORS" WO 99/21979; Del Cardayre et al. "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" WO 98/1837; Patten and Stemmer, "METHODS AND COMPOSITIONS FOR POLYPEPTIDE ENGINEERING" WO 98/2720; Stemmer et al., and "METHODS FOR OPTIMIZATION OF GENE THERAPY BY RECURSIVE SEQUENCE SHUFFLING AND SELECTION" WO9813487.

Certain U.S. Applications provide additional details regarding DNA shuffling and related techniques, as well as other diversity generating methods,

including "SHUFFLING OF CODON ALTERED GENES" by Pattem et al. filed and September 28, 1999; USSN 09/407,800; "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION", by del Cardyre et al. filed July 15, 1999 (USSN 09/354,922); "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed January 18, 2000 (PCT/US00/01202); "USE OF CODON-BASED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., filed September 28, 1999 (USSN 09/408,393); and "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Seifonov et al., USSN 09/484,850 and PCT/US00/01203, filed January 18, 2000.

As review of the foregoing publications, patents, published applications and U.S. patent applications reveals, recursive recombination of nucleic acids to provide new nucleic acids with desired properties can be carried out by a number of established methods and these procedures can be combined with any of a variety of other diversity generating methods.

In brief, at least 5 different general classes of recombination methods are applicable to the present invention and set forth in the references above. First, nucleic acids can be recombined *in vitro* by any of a variety of techniques discussed in the references above, including e.g., DNase digestion of nucleic acids to be recombined followed by ligation and/or PCR reassembly of the nucleic acids. Second, nucleic acids can be recursively recombined *in vivo*, e.g., by allowing recombination to occur between nucleic acids in cells. Third, whole genome recombination methods can be used in which whole genomes (or significant fractions thereof) of cells or other organisms are recombined, optionally including splicing of the genomic recombination mixtures with selected nucleic acids (e.g., which encode lipid synthesis enzymes or other relevant factors for enhanced lipid biosynthesis, as noted herein). Fourth, synthetic recombination methods can be used, in which oligonucleotides corresponding to targets of interest are synthesized and reassembled in PCR and/or ligation reactions which include oligonucleotides which correspond to more than one parental nucleic acid (e.g., including one or more lipid biosynthetic nucleic acid), thereby generating new recombined nucleic acids. Oligonucleotides can be made by standard nucleotide addition methods, or can be made, e.g., by tri-nucleotide synthetic approaches. Fifth, in

silico methods of recombination can be effected in which genetic algorithms are used in a computer to recombine sequence strings which correspond to nucleic acid homologues (or even non-homologous) sequences. The resulting recombined sequence strings are optionally converted into nucleic acids by synthesis of nucleic acids which correspond to the recombined sequences, e.g., in concert with oligonucleotide synthesis/ gene reassembly techniques. Any of the preceding general recombination formats can be practiced in a reiterative fashion to generate a more diverse set of recombinant nucleic acids. The methods can also be practiced in combination.

The above references provide these and other basic recombination formats as well as many modifications of these formats. Regardless of the format which is used, the nucleic acids of the invention can be recombined (with each other or with related (or even unrelated) nucleic acids to produce a diverse set of recombinant nucleic acids, including, e.g., sets of homologous or non-homologous nucleic acids.

Following recombination, any nucleic acids which are produced can be selected for a desired activity. In the context of the present invention, this can include testing for and identifying any activity that can be detected, including in an automatable format, by any of the assays in the art. A variety of related (or even unrelated) properties can be assayed for, using any available assay.

DNA shuffling and related techniques provide a robust, widely applicable, means of generating diversity useful for the engineering of proteins, pathways, cells and organisms with improved characteristics. In addition to the basic formats described above, it is sometimes desirable to combine recombination methodologies with additional techniques for generating diversity. In conjunction with (or separately from) recombination-based methods, a variety of other diversity generation methods can be practiced and the results (i.e., diverse populations of nucleic acids) screened for. Additional diversity can be introduced into nucleic acids by methods which result in the alteration of individual nucleotides or groups of contiguous or non-contiguous nucleotides, e.g., mutagenesis methods. Mutagenesis methods include, for example, recombination (PCT/US98/05222; Publ. No. WO98/42727), oligonucleotide-directed mutagenesis (for review *see*, Smith, *Ann. Rev. Genet.* 19: 423-462 (1985)); Bolstein and Shortle, *Science* 229: 1193-1201 (1983); Carter, *Biochem. J.* 237: 1-7 (1986); Kunkel, "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic acids & Molecular Biology*, Eckstein and Lilley, eds., Springer Verlag, Berlin

- (1987). Included among these methods are oligonucleotide-directed mutagenesis (Zoller and Smith, *Nucl. Acids Res.* 10: 6487-6500 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 329-350 (1987)) phosphothioate-modified DNA mutagenesis (Taylor et al., *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor et al., *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye and Eckstein, *Nucl. Acids Res.* 14: 9679-9698 (1986); Sayers et al., *Nucl. Acids Res.* 16: 791-802 (1988); Sayers et al., *Nucl. Acids Res.* 16: 803-814 (1988)), mutagenesis using uracil-containing templates (Kunkel, *Proc. Natl. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel et al., *Methods in Enzymol.* 154: 367-382); mutagenesis using gapped duplex DNA (Kramer et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer and Fritz, *Methods in Enzymol.* 154: 350-367 (1987); Kramer et al., *Nucl. Acids Res.* 16: 7207 (1988); and Fritz et al., *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional suitable methods include point mismatch repair (Kramer et al., *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Carter et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1987)), deletion mutagenesis (Eghtedarzadeh and Henikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction-selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakamar and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grundstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1985)). Kits for mutagenesis are commercially available (e.g., Bio-Rad, Amersham International, Anglian Biotechnology).
- Other relevant references which describe methods of diversity nucleic acids include Schellenger U.S. Patent No. 5,756,316; U.S. Patent No. 5,965,408; Ostermeier et al. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" *Nature Biotech* 17:1205; U.S. Patent No. 5,783,431; U.S. Patent No. 5,824,485; U.S. Patent 5,958,672; Jitholt et al. (1998) "Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework" *Gene* 215: 471; U.S. Patent No. 5,939,250; WO 99/10539; WO 98/88085 and WO 99/10539.
- Any of these diversity generating methods can be combined, in any combination selected by the user, to produce nucleic acid diversity, which may be screened for using any available screening method.

#### CREATION OF RECOMBINANT LIBRARIES

The invention involves creating recombinant libraries of polynucleotides that are then screened to identify those library members that exhibit a desired property, e.g., which encode lipid synthetic or metabolic activity. The recombinant libraries can be created using any of the various methods herein, as well as many others which would be apparent to one of skill. Methods for obtaining recombinant polynucleotides and/or for obtaining diversity in nucleic acids used as the substrates for DNA shuffling as described herein include, for example, those references noted in the preceding section, including those related to recombination, mutation and other diversity generation procedures.

In a presently preferred embodiment, the recombinant libraries are prepared, at least in part, using DNA shuffling. Iterative cycles of recombination and screening/selection can be performed to further evolve any nucleic acid(s) of interest. Such techniques do not require the extensive analysis and computation required by conventional methods for polypeptide engineering. Shuffling allows the recombination of large numbers of mutations in a minimum number of selection cycles, in contrast to traditional, pairwise recombination events. Thus, the sequence recombination techniques described herein provide particular advantages in that they provide recombination between mutations in any or all of these, thereby providing a very fast way of exploring the manner in which different combinations of mutations can effect a desired result. In some instances, however, structural and/or functional information is available which, although not required for sequence recombination, provides opportunities for modification of the technique.

#### ADDITIONAL SHUFFLING FORMAT DETAILS

The breeding procedure starts with at least two substrates that generally show substantial sequence identity to each other (i.e., at least about 30%, 50%, 70%, 80% or 90% sequence identity), but differ from each other at certain positions. The difference can be any type of mutation, for example, substitutions, insertions, deletions or the like. Often, e.g., in *in vitro* shuffling procedures, different segments differ from each other in about 5-20 positions. For physical recombination to generate increased diversity relative to the starting materials, the starting materials typically differ from each other in at least two nucleotide positions. That is, if there are only two substrates,

there are usually at least two divergent positions. If there are three substrates, for example, one substrate can differ from the second at a single position, and the second can differ from the third at a different single position. The starting DNA segments can be natural variants of each other, for example, allelic or species variants. The segments can also be from nonallelic genes showing some degree of structural and usually (though not necessarily) functional relatedness (e.g., different genes within a superfamily).

The starting DNA segments can also be induced variants of each other. For example, one DNA segment can be produced by error-prone PCR replication of the other, or by substitution of a mutagenic cassette. Induced mutants can also be prepared by propagating one (or both) of the segments in a mutagenic strain. In these situations, the second DNA segment is not generally a single segment but a large family of related segments. The different segments forming the starting materials are often the same length or substantially the same length. However, this need not be the case; for example, one segment can be a subsequence of another. The segments can be present as part of larger molecules, such as vectors, or can be in isolated form.

The starting DNA segments are recombined by any of the sequence recombination formats provided herein to generate a diverse library of recombinant DNA segments. Such a library can vary widely in size from having fewer than 10 to more than  $10^5$ ,  $10^6$ ,  $10^7$  or more members. In some embodiments, the starting segments and the recombinant libraries generated will include full-length coding sequences and any essential regulatory sequences, such as a promoter and polyadenylation sequence, required for expression. In other embodiments, the recombinant DNA segments in the library can be inserted into a common vector providing sequences necessary for expression before performing screening/selection.

#### Use of Restriction Enzyme Sites to Recombine Mutations

In some situations it is advantageous to use restriction enzyme sites in nucleic acids to direct the recombination of mutations in a nucleic acid sequence of interest. These techniques are particularly preferred in the evolution of fragments that cannot readily be shuffled by existing methods due to the presence of repeated DNA or other problematic primary sequence motifs. These situations also include recombination formats in which it is preferred to retain certain sequences unmutated. The use of restriction enzyme sites is also preferred for shuffling large fragments

(typically greater than 10 kb), such as gene clusters that cannot be readily shuffled and "PCR-amplified" because of their size. Although fragments up to 50 kb have been reported to be amplified by PCR (Barnes, *Proc. Natl. Acad. Sci. U.S.A.* 91:2216-2220 (1994)), it can be problematic for fragments over 10 kb, and thus alternative methods for shuffling in the range of 10 - 50 kb and beyond are preferred. Preferably, the restriction endonucleases used are of the Class II type (Sambrook, Ausubel and Berger, *supra*) and of these, preferably those which generate nonpalindromic sticky end overhangs such as AlwI, Sfi I or BstXI. These enzymes generate nonpalindromic ends that allow for efficient ordered reassembly with DNA ligase. Typically, restriction enzyme (or endonuclease) sites are identified by conventional restriction enzyme mapping techniques (Sambrook, Ausubel, and Berger, *supra*), by analysis of sequence information for that gene, or by introduction of desired restriction sites into a nucleic acid sequence by synthesis (i.e. by incorporation of silent mutations).

The DNA substrate molecules to be digested can either be from *in vivo* replicated DNA, such as a plasmid preparation, or from PCR amplified nucleic acid fragments harboring the restriction enzyme recognition sites of interest, preferably near the ends of the fragment. Typically, at least two variants of a gene of interest, each having one or more mutations, are digested with at least one restriction enzyme determined to cut within the nucleic acid sequence of interest. The restriction fragments are then joined with DNA ligase to generate full length genes having shuffled regions. The number of regions shuffled will depend on the number of cuts within the nucleic acid sequence of interest. The shuffled molecules can be introduced into cells as described above and screened or selected for a desired property as described herein. Nucleic acid can then be isolated from pools (libraries), or clones having desired properties and subjected to the same procedure until a desired degree of improvement is obtained.

In some embodiments, at least one DNA substrate molecule or fragment thereof is isolated and subjected to mutagenesis. In some embodiments, the pool or library of religated restriction fragments are subjected to mutagenesis before the digestion-ligation process is repeated. "Mutagenesis" as used herein comprises such techniques known in the art as PCR mutagenesis, oligonucleotide-directed mutagenesis, site-directed mutagenesis, etc., and recursive sequence recombination by any of the techniques described herein.

### Reassembly PCR

A further technique for recombining mutations in a nucleic acid sequence utilizes "reassembly PCR." This method can be used to assemble multiple segments that have been separately evolved into a full length nucleic acid template such as a gene. This technique is performed when a pool of advantageous mutants is known from previous work or has been identified by screening mutants that may have been created by any mutagenesis technique known in the art, such as PCR mutagenesis, cassette mutagenesis, doped oligo mutagenesis, chemical mutagenesis, or propagation of the DNA template *in vivo* in mutator strains. Boundaries defining segments of a nucleic acid sequence of interest preferably lie in intergenic regions, introns, or areas of a gene not likely to have mutations of interest. Preferably, oligonucleotide primers (oligos) are synthesized for PCR amplification of segments of the nucleic acid sequence of interest, such that the sequences of the oligonucleotides overlap the junctions of two segments. The overlap region is typically about 10 to 100 nucleotides in length. Each of the segments is amplified with a set of such primers. The PCR products are then "reassembled" according to assembly protocols such as those discussed herein to assemble randomly fragmented genes. In brief, in an assembly protocol the PCR products are first purified away from the primers, by, for example, gel electrophoresis or size exclusion chromatography. Purified products are mixed together and subjected to about 1-10 cycles of denaturing, reannealing, and extension in the presence of polymerase and deoxynucleoside triphosphates (dNTP's) and appropriate buffer salts in the absence of additional primers ("self-priming"). Subsequent PCR with primers flanking the gene are used to amplify the yield of the fully reassembled and shuffled genes.

In some embodiments, the resulting reassembled genes are subjected to mutagenesis before the process is repeated.

In a further embodiment, the PCR primers for amplification of segments of the nucleic acid sequence of interest are used to introduce variation into the gene of interest as follows. Mutations at sites of interest in a nucleic acid sequence are identified by screening or selection, by sequencing homologues of the nucleic acid sequence, and so on. Oligonucleotide PCR primers are then synthesized which encode wild type or mutant information at sites of interest. These primers are then used in PCR mutagenesis to generate libraries of full length genes encoding permutations of wild

type and mutant information at the designated positions. This technique is typically advantageous in cases where the screening or selection process is expensive, cumbersome, or impractical relative to the cost of sequencing the genes of mutants of interest and synthesizing mutagenic oligonucleotides.

### Site Directed Mutagenesis (SDM) with Oligonucleotides Encoding Homologue Mutations Followed by Shuffling

In some embodiments of the invention, sequence information from one or more substrate sequences is added to a given "parental" sequence of interest, with subsequent recombination between rounds of screening or selection. Typically, this is done with site-directed mutagenesis performed by techniques well known in the art (e.g., Berger, Ausubel and Sambrook, *supra*) with one substrate as template and oligonucleotides encoding single or multiple mutations from other substrate sequences, e.g. homologue genes. After screening or selection for an improved phenotype of interest, the selected recombinant(s) can be further evolved using HSR techniques described herein. After screening or selection, site-directed mutagenesis can be done again with another collection of oligonucleotides encoding homologue mutations, and the above process repeated until the desired properties are obtained.

When the difference between two homologues is one or more single point mutations in a codon, degenerate oligonucleotides can be used that encode the sequences in both homologues. One oligonucleotide can include many such degenerate codons and still allow one to exhaustively search all permutations over that block of sequence.

When the homologue sequence space is very large, it can be advantageous to restrict the search to certain variants. Thus, for example, computer modeling tools (Lathrop *et al.* (1996) *J. Mol. Biol.*, 255: 641-665) can be used to model each homologue mutation onto the target protein and discard any mutations that are predicted to grossly disrupt structure and function.

### In Vitro DNA Shuffling Formats

In one embodiment for shuffling DNA sequences *in vitro*, the initial substrates for recombination are a pool of related sequences, e.g., different, variant forms, as homologs from different individuals, strains, or species of an organism, or related sequences from the same organism, as allelic variations. The sequences can be DNA or RNA and can be of various lengths depending on the size of the gene or DNA

fragment to be recombined or reassembled. Preferably the sequences are from about 50 base pairs (bp) to about 50 kilobases (kb).

The pool of related substrates are converted into overlapping fragments, *e.g.*, from about 5 bp to 5 kb or more. Often, for example, the size of the fragments is from about 10 bp to 1000 bp, and sometimes the size of the DNA fragments is from about 100 bp to 500 bp. The conversion can be effected by a number of different methods, such as DNase I or RNase digestion, random shearing or partial restriction enzyme digestion. For discussions of protocols for the isolation, manipulation, enzymatic digestion, and the like of nucleic acids, see, for example, Sambrook *et al.*, and Ausubel, both *supra*. The concentration of nucleic acid fragments of a particular length and sequence is often less than 0.1 % or 1% by weight of the total nucleic acid. The number of different specific nucleic acid fragments in the mixture is usually at least about 100, 500 or 1000.

The mixed population of nucleic acid fragments are converted to at least partially single-stranded form using a variety of techniques, including, for example, heating, chemical denaturation, use of DNA binding proteins, and the like. Conversion can be effected by heating to about 80°C to 100°C, more preferably from 90°C to 96°C, to form single-stranded nucleic acid fragments and then reannealing. Conversion can also be effected by treatment with single-stranded DNA binding protein (see Wold (1997) *Annu. Rev. Biochem.* 66:61-92) or *recA* protein (see, *e.g.*, Kilianisa (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:7837-7840). Single-stranded nucleic acid fragments having regions of sequence identity with other single-stranded nucleic acid fragments can then be reannealed by cooling to 20°C to 75°C, and preferably from 40°C to 65°C.

Renaturation can be accelerated by the addition of polyethylene glycol (PEG), other volume-excluding reagents or salt. The salt concentration is preferably from 0 mM to 200 mM, more preferably the salt concentration is from 10 mM to 100 mM. The salt may be KCl or NaCl. The concentration of PEG is preferably from 0% to 20%, more preferably from 5% to 10%. The fragments that reanneal can be from different substrates. The annealed nucleic acid fragments are incubated in the presence of a nucleic acid polymerase, such as Taq or Klenow, and dNTP's (*i.e.* dATP, dCTP, dGTP and dTTP). If regions of sequence identity are large, Taq polymerase can be used with an annealing temperature of between 45-65°C. If the areas of identity are small,

Klenow polymerase can be used with an annealing temperature of between 20-30°C. The polymerase can be added to the random nucleic acid fragments prior to annealing, simultaneously with annealing or after annealing.

The process of denaturation, renaturation and incubation in the presence of polymerase of overlapping fragments to generate a collection of polynucleotides containing different permutations of fragments is sometimes referred to as shuffling of the nucleic acid *in vitro*. This cycle is repeated for a desired number of times. Preferably the cycle is repeated from 2 to 100 times, more preferably the sequence is repeated from 10 to 40 times. The resulting nucleic acids are a family of double-stranded polynucleotides of from about 50 bp to about 100 kb, preferably from 500 bp to 50 kb. The population represents variants of the starting substrates showing substantial sequence identity thereto but also diverging at several positions. The population has many more members than the starting substrates. The population of fragments resulting from shuffling is used to transform host cells, optionally after cloning into a vector.

In one embodiment utilizing *in vitro* shuffling, subsequences of recombination substrates can be generated by amplifying the full-length sequences under conditions which produce a substantial fraction, typically at least 20 percent or more, of incompletely extended amplification products. Another embodiment uses random primers to prime the entire template DNA to generate less than full length amplification products. The amplification products, including the incompletely extended amplification products are denatured and subjected to at least one additional cycle of reannealing and amplification. This variation, in which at least one cycle of reannealing and amplification provides a substantial fraction of incompletely extended products, is termed "snitching." In the subsequent amplification round, the partially extended (less than full length) products reanneal to and prime extension on different sequence-related template species. In another embodiment, the conversion of substrates to fragments can be effected by partial PCR amplification of substrates.

In another embodiment, a mixture of fragments is spiked with one or more oligonucleotides. The oligonucleotides can be designed to include precharacterized mutations of a wildtype sequence, or sites of natural variations between individuals or species. The oligonucleotides also include sufficient sequence or structural homology flanking such mutations or variations to allow annealing with

the wildtype fragments. Annealing temperatures can be adjusted depending on the length of homology.

- In a further embodiment, recombination occurs in at least one cycle by template switching, such as when a DNA fragment derived from one template primes on the homologous position of a related but different template. Template switching can be induced by addition of recA (see, Kianitisa (1997) *supra*), rad51 (see, Namsarev (1997) *Mol. Cell. Biol.* 17:5359-5368), rad55 (see, Clever (1997) *EMBO J.* 16:2535-2544), rad57 (see, Sung (1997) *Genes Dev.* 11:1111-1121) or other polymerases (e.g., viral polymerases, reverse transcriptase) to the amplification mixture.
- Template switching can also be increased by increasing the DNA template concentration.

- Another embodiment utilizes at least one cycle of amplification, which can be conducted using a collection of overlapping single-stranded DNA fragments of related sequence, and different lengths. Fragments can be prepared using a single stranded DNA phage, such as M13 (see, Wang (1997) *Biochemistry* 36:9486-9492). Each fragment can hybridize to and prime polynucleotide chain extension of a second fragment from the collection, thus forming sequence-recombined polynucleotides. In a further variation, ssDNA fragments of variable length can be generated from a single primer by Pfu, Taq, Vent, Deep Vent, UTMa DNA polymerase or other DNA polymerases on a first DNA template (see, Cline (1996) *Nucleic Acids Res.* 24:3546-3551). The single stranded DNA fragments are used as primers for a second, Kunkel-type template, consisting of a uracil-containing circular ssDNA. This results in multiple substitutions of the first template into the second. See, Levichkin (1995) *Mol. Biology* 29:572-577; Jung (1992) *Gene* 121:17-24.

- In some embodiments of the invention, shuffled nucleic acids obtained by use of the recursive recombination methods of the invention, are put into a cell and/or organism for screening. Shuffled lipid synthetic genes can be introduced into, for example, bacterial cells, yeast cells, or plant cells for initial screening. *Bacillus* species (such as *B. subtilis* and *E. coli* are two examples of suitable bacterial cells into which one can insert and express shuffled genes. The shuffled genes can be introduced into bacterial or yeast cells either by integration into the chromosomal DNA or as plasmids. Shuffled genes can also be introduced into plant cells for screening purposes. Thus, a transgene of interest can be modified using the recursive sequence

recombination methods of the invention *in vitro* and reinserted into the cell for *in vivo/in situ* selection for the new or improved property.

#### In Vivo DNA Shuffling Formats

- In some embodiments of the invention, DNA substrate molecules are introduced into cells, wherein the cellular machinery directs their recombination. For example, a library of mutants is constructed and screened or selected for mutants with improved phenotypes by any of the techniques described herein. The DNA substrate molecules encoding the best candidates are recovered by any of the techniques described herein, then fragmented and used to transfect a plant host and screened or selected for improved function. If further improvement is desired, the DNA substrate molecules are recovered from the plant host cell, such as by PCR, and the process is repeated until a desired level of improvement is obtained. In some embodiments, the fragments are denatured and reannealed prior to transfection, coated with recombination stimulating proteins such as recA, or co-transfected with a selectable marker such as Neo<sup>R</sup> to allow the positive selection for cells receiving recombined versions of the gene of interest. Methods for *in vivo* shuffling are described in, for example, PCT applications WO 98/13487 and WO 97/ 07205.

- The efficiency of *in vivo* shuffling can be enhanced by increasing the copy number of a gene of interest in the host cells. For example, the majority of bacterial cells in stationary phase cultures grown in rich media contain two, four or eight genomes. In minimal medium the cells contain one or two genomes. The number of genomes per bacterial cell thus depends on the growth rate of the cell as it enters stationary phase. This is because rapidly growing cells contain multiple replication forks, resulting in several genomes in the cells after termination. The number of genomes is strain dependent, although all strains tested have more than one chromosome in stationary phase. The number of genomes in stationary phase cells decreases with time. This appears to be due to fragmentation and degradation of entire chromosomes, similar to apoptosis in mammalian cells. This fragmentation of genomes in cells containing multiple genome copies results in massive recombination and mutagenesis. The presence of multiple genome copies in such cells results in a higher frequency of homologous recombination in these cells, both between copies of a gene in different genomes within the cell, and between a genome within the cell and a

transfected fragment. The increased frequency of recombination allows one to evolve a gene evolved more quickly to acquire optimized characteristics.

In nature, the existence of multiple genomic copies in a cell type would usually not be advantageous due to the greater nutritional requirements needed to maintain this copy number. However, artificial conditions can be devised to select for high copy number. Modified cells having recombinant genomes are grown in rich media (in which conditions, multicopy number should not be a disadvantage) and exposed to a mutagen, such as ultraviolet or gamma irradiation or a chemical mutagen, e.g., mitomycin, nitrous acid, photoactivated psoralens, alone or in combination, which induces DNA breaks amenable to repair by recombination. These conditions select for cells having multicopy number due to the greater efficiency with which mutations can be excised. Modified cells surviving exposure to mutagen are enriched for cells with multiple genome copies. If desired, selected cells can be individually analyzed for genome copy number (e.g., by quantitative hybridization with appropriate controls). For example, individual cells can be sorted using a cell sorter for those cells containing more DNA, e.g., using DNA specific fluorescent compounds or sorting for increased size using light dispersion. Some or all of the collection of cells surviving selection are tested for the presence of a gene that is optimized for the desired property.

In one embodiment, phage libraries are made and recombined in mutator strains such as cells with mutant or impaired gene products of *mutS*, *mutT*, *mutH*, *mutL*, *ovrD*, *dam*, *var*, *umuD*, *spoB*, *recI*, etc. The impairment is achieved by genetic mutation, allelic replacement, selective inhibition by an added reagent such as a small compound or an expressed antisense RNA, or other techniques. High multiplicity of infection (MOI) libraries are used to infect the cells to increase recombination frequency.

Additional strategies for making phage libraries and/or for recombining DNA from donor and recipient cells are set forth in U.S. Pat. No. 5,521,077. Additional recombination strategies for recombining plasmids in yeast are set forth in WO 97 07205.

### 30 Whole Genome Shuffling

In one embodiment, the selection methods herein are utilized in a "whole genome shuffling" format. An extensive guide to the many forms of whole genome shuffling is found in the pioneering application to the inventors and their co-workers,

e.g., Del Cardayre et al. "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" WO 9831837 and "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION", by del Cardayre et al. filed July 15, 1999 (USSN 09/354,922).

In brief, whole genome shuffling makes no presuppositions at all regarding what nucleic acids may confer a desired property. Instead, entire genomes (e.g., from a genomic library, or isolated from an organism) are shuffled in cells and selection protocols applied to the cells.

An application of recursive whole genome shuffling is the evolution of plant cells, and transgenic plants derived from the same, to acquire desirable lipid production properties. The substrates for recombination can be, e.g., whole genomic libraries, fractions thereof or focused libraries containing variants of gene(s) known or suspected to confer tolerance to one of the above agents. Frequently, library fragments are obtained from a different species to the plant being evolved. Regardless of the precise shuffling methodology used, the selection methods described above for lipid biosynthetic selection, including selection for any of the desirable traits noted herein can be performed.

The DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (From et al., *Proc. Natl. Acad. Sci. USA* 82, 5824 (1985), infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., *Molecular Biology of Plant Tumors*, (Academic Press, New York, 1982) pp. 549-560; Howell, US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., *Nature* 327, 70-73 (1987)), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al., *Science* 233, 496-498 (1984); Fraley et al., *Proc. Natl. Acad. Sci. USA* 80, 4803 (1983)).

Diversity can also be generated by genetic exchange between plant protoplasts. Procedures for formation and fusion of plant protoplasts are described by Takahashi et al., US 4,677,066; Akagi et al., US 5,360,725; Shimamoto et al., US 5,250,433; Cheney et al., US 5,426,040.

After a suitable period of incubation to allow recombination to occur and for expression of recombinant genes, the plant cells are assayed for lipid production (e.g., membrane lipid composition), and suitable plant cells are collected. Some or all of these plant cells can be subject to a further round of recombination and screening. Eventually, plant cells having the required degree of lipid expression are obtained. This is especially suitable, e.g., for plant oil accumulation in seeds.

These cells can then be cultured into transgenic plants. Plant regeneration from cultured protoplasts is described in Evans et al., "Protoplast Isolation and Culture," *Handbook of Plant Cell Cultures* 1, 124-176 (MacMillan Publishing Co., New York, 1983); Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," *Protoplasts*, (1983) pp. 12-29, (Birkhauser, Basel 1983); Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," *Protoplasts* (1983) pp. 31-41, (Birkhauser, Basel 1983); Binding, "Regeneration of Plants," *Plant Protoplasts*, pp. 21-73, (CRC Press, Boca Raton, 1985) and other references available to persons of skill. Additional details regarding plant regeneration from cells are also found below.

In a variation of the above method, one or more preliminary rounds of recombination and screening can be performed in bacterial cells according to the same general strategy as described for plant cells. More rapid evolution can be achieved in bacterial cells due to their greater growth rate and the greater efficiency with which DNA can be introduced into such cells. After one or more rounds of recombination/screening, a DNA fragment library is recovered from bacteria and transformed into the plants. The library can either be a complete library or a focused library. A focused library can be produced by amplification from primers specific for plant sequences, particularly plant sequences known or suspected to have a role in conferring a desirable lipid production or metabolic property.

Plant genome shuffling allows recursive cycles to be used for the introduction and recombination of genes or pathways that confer improved properties to desired plant species. Any plant species, including weeds and wild cultivars, showing a desired trait, such as high oil production, can be used as the source of DNA that is introduced into the crop or horticultural host plant species.

Genomic DNA prepared from the source plant is fragmented (e.g., by DNaseI, restriction enzymes, or mechanically) and cloned into a vector suitable for

making plant genomic libraries, such as pGA482 (An. G., 1995, *Methods Mol. Biol.* 44:47-58). This vector contains the *A. tumefaciens* left and right borders needed for gene transfer to plant cells and antibiotic markers for selection in *E. coli*, *Agrobacterium*, and plant cells. A multicloning site is provided for insertion of the genomic fragments. A cos sequence is present for the efficient packaging of DNA into bacteriophage lambda heads for transfection of the primary library into *E. coli*. The vector accepts DNA fragments of 25-40 kb.

The primary library can also be directly electroporated into an *A. tumefaciens* or *A. rhizogenes* strain that is used to infect and transform host plant cells (Main, GD et al., 1995, *Methods Mol. Biol.* 44:405-412). Alternatively, DNA can be introduced by electroporation or PEG-mediated uptake into protoplasts of the recipient plant species (Bilang et al. (1994) *Plant Mol. Biol. Manual*, Kluwer Academic Publishers, A1:1-16) or by particle bombardment of cells or tissues (Christou, Tolid, A2:1-15). If necessary, antibiotic markers in the T-DNA region can be eliminated, as long as selection for the trait is possible, so that the final plant products contain no antibiotic genes.

Stably transformed whole cells acquiring the trait are selected on solid or liquid media. If the trait in question cannot be selected for directly, transformed cells can be selected with antibiotics and allowed to form callus or regenerated to whole plants and then screened for the desired property.

The second and further cycles consist of isolating genomic DNA from each transgenic line and introducing it into one or more of the other transgenic lines. In each round, transformed cells are selected or screened, typically in an incremental fashion (increasing dosages, etc.). To speed the process of using multiple cycles of transformation, plant regeneration can be eliminated until the last round. Callus tissue generated from the protoplasts or transformed tissues can serve as a source of genomic DNA and new host cells. After the final round, fertile plants are regenerated and the progeny are selected for homozygosity of the inserted DNAs. Ultimately, a new plant is created that carries multiple inserts which additively or synergistically combine to confer high levels of the desired trait.

In addition, the introduced DNA that confers the desired trait can be traced because it is flanked by known sequences in the vector. Either PCR or plasmid rescue is used to isolate the sequences and characterize them in more detail. Long PCR

(Food, OS and Rose, E.A. 1995, PCR Primer, A Laboratory Manual, CSHL Press, pp 63-77) of the full 25-40 kb insert is achieved with the proper reagents and techniques using as primers the T-DNA border sequences. If the vector is modified to contain the *E. coli* origin of replication and an antibiotic marker between the T-DNA borders, a rare cutting restriction enzyme, such as NotI or SfiI, that cuts only at the ends of the inserted DNA is used to create fragments containing the source plant DNA that are then self-ligated and transformed into *E. coli* where they replicate as plasmids. The total DNA or subfragment of it that is responsible for the transferred trait can be subjected to in vitro evolution by DNA shuffling. The shuffled library is then introduced into host plant cells and screened for improvement of the trait. In this way, single and multigene traits can be transferred from one species to another and optimized for higher expression or activity leading to whole organism improvement.

#### Oligonucleotide and in silico shuffling formats

In addition to the formats for shuffling noted above, at least two additional related formats are useful in the practice of the present invention. The first, referred to as "in silico" shuffling utilizes computer algorithms to perform virtual shuffling using genetic operators in a computer. As applied to the present invention, lipid synthetic or metabolic gene sequence strings (or sequence strings corresponding to any other factors which affect lipid biosynthesis or metabolism) are recombined in a computer system and desirable products are made, e.g., by reassembly PCR of synthetic oligonucleotides. In silico shuffling is described in detail in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., USSN 09/484,850 and PCT/US00/01203, filed January 18, 2000.

The second useful format is referred to as "oligonucleotide mediated shuffling" in which oligonucleotides corresponding to a family of related homologous nucleic acids (e.g., as applied to the present invention, interspecific or allelic variants of a lipid metabolic or synthetic nucleic acid) are recombined to produce selectable nucleic acids. This format is described in detail, e.g., in "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed January 18, 2000 (PCT/US00/01202) and in "USE OF CODON-BASED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., filed September 28, 1999 (USSN 09/408,393).

#### MAKING TRANGENIC PLANTS

In one aspect, nucleic acids shuffled for lipid synthetic or metabolic activity by any of the techniques noted above are used to make transgenic plants, thereby providing transgenic plants. Methods of transducing plant cells with nucleic acids are generally available. In addition to Berger, Ausubel and Sambrook, useful general references for plant cell cloning, culture and regeneration include Payne et al. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, NY (Payne) and Gamborg and Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture: Fundamental Methods* Springer Lab Manual, Springer-Verlag (Berlin).  
Heideberg New York (Gamborg). Cell culture media are described in Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL (Atlas). Additional information is found in commercial literature such as the Life Science Research Cell Culture catalogue (1998) from Sigma-Aldrich, Inc. (St Louis, MO) (Sigma-LSRCCC) and, e.g., the Plant Culture Catalogue and supplement (1997) also from Sigma-Aldrich, Inc. (St Louis, MO) (Sigma-PCCS).

In one embodiment of this invention, recombinant DNA vectors which contain isolated selected shuffled sequences and are suitable for transformation of plant cells are prepared. A DNA sequence coding for the desired nucleic acid, for example a cDNA or a genomic sequence encoding a full length protein, is conveniently used to construct a recombinant expression cassette which can be introduced into the desired plant. An expression cassette will typically comprise a selected shuffled nucleic acid sequence operably linked to a promoter sequence and other transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues (e.g., entire plant, leaves, seeds) of the transformed plant.

For example, a strongly or weakly constitutive plant promoter can be employed which will direct expression of a shuffled enzyme gene as set forth herein in all tissues of a plant. Such promoters are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill. Where overexpression of a shuffled gene is detrimental to the plant, one of skill, upon review of this disclosure, will recognize that weak constitutive promoters can be used

for low levels of expression. In those cases where high levels of expression is not harmful to the plant, a strong promoter, e.g., a t-RNA or other pol III promoter, or a strong pol II promoter, such as the cauliflower mosaic virus promoter, can be used.

Alternatively, a plant promoter may be under environmental control.

Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light.

In one embodiment of this invention, the promoters used in the constructs of the invention will be "tissue-specific" and are under developmental control such that the desired gene is expressed only in certain tissues, such as leaves and seeds. The endogenous promoters (or shuffled variants thereof) from lipid synthetic genes are particularly useful for directing expression of these genes to the transfecting plant.

Tissue-specific promoters can also be used to direct expression of heterologous structural genes, including shuffled nucleic acids as described herein. Examples include genes encoding proteins which ordinarily provide the plant with lipid synthetic activity and genes that encode useful phenotypic characteristics, e.g., which influence heterosis.

In general, the particular promoter used in the expression cassette in plants depends on the intended application. Any of a number of promoters which direct transcription in plant cells can be suitable. The promoter can be either constitutive or inducible. In addition to the promoters noted above, promoters of bacterial origin which operate in plants include the octopine synthase promoter, the nopaline synthase promoter and other promoters derived from native Ti plasmids. See, Herrera-Estrella *et al.* (1983), *Nature*, 303:209-213. Viral promoters include the 35S and 19S RNA promoters of cauliflower mosaic virus. See, Odell *et al.* (1985), *Nature*, 313:810-812. Other plant promoters include the ribulose-1,3-bisphosphate carboxylase small subunit promoter and the phased-in promoter. The promoter sequence from the E8 gene and other genes may also be used. The isolation and sequence of the E8 promoter is described in detail in Dekman and Fischer, (1988) *EMBO J.* 7:3315-3327.

To identify candidate promoters, the 5' portions of a genomic clone is analyzed for sequences characteristic of promoter sequences. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAAT), which is

usually 20 to 30 base pairs upstream of the transcription start site. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. Messing *et al.*, GENETIC ENGINEERING IN PLANTS, Kosuge, *et al.* (eds.), pp. 221-227 (1983).

In preparing expression vectors of the invention, sequences other than the promoter and the shuffled gene are also preferably used. If normal polypeptide expression is desired, a polyadenylation region at the 3'-end of the shuffled coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the shuffled sequence will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide tolerance, particularly antibiotic tolerance, such as tolerance to kanamycin, G418, bleomycin, hygromycin, or herbicide tolerance, such as tolerance to chlorosulfuron, or phosphinothricin (the active ingredient in the herbicides bialaphos and Basta).

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, e.g., Payne, Gamberg, Atlas, Sigma-LSRCCC and Sigma-PCCS, *all supra*, as well as, e.g., Weising, *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988).

For example, DNAs may be introduced directly into the genomic DNA of a plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, *et al.*, *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm, *et al.*, *Proc. Natl. Acad. Sci. USA*

82:5824 (1985). Ballistic transformation techniques are described in Klein, *et al.*, *Nature* 327:70-73 (1987); and Weeks, *et al.*, *Plant Physiol.* 102:1077-1084 (1993).

In a particularly preferred embodiment, *Agrobacterium tumefaciens*-mediated transformation techniques are used to transfer shuffled coding sequences to transgenic plants. *Agrobacterium*-mediated transformation is useful primarily in dicots, however, certain monocots can be transformed by *Agrobacterium*. For instance, *Agrobacterium* transformation of rice is described by Hiei, *et al.*, *Plant J.* 6:271-282 (1994); U.S. Patent No. 5,187, 073; U.S. Patent 5,591,616; Li, *et al.*, *Science in China* 34:54 (1991), and Raimaci, *et al.*, *BioTechnology* 8:33 (1990). Xu, *et al.*, *Chinese J. Bot.* 2:81 (1990) transformed maize, barley, triticale and asparagus by *Agrobacterium* infection.

In this technique, the ability of the tumour-inducing (Ti) plasmid of *A. tumefaciens* to integrate into a plant cell genome is used advantageously to co-transfer a nucleic acid of interest into a recombinant plant cell of this invention. Typically, an expression vector is produced wherein the nucleic acid of interest is ligated into an autonomously replicating plasmid which also contains T-DNA sequences. T-DNA sequences typically flank the expression cassette nucleic acid of interest and comprise the integration sequences of the plasmid. In addition to the expression cassette, T-DNA also typically comprises a marker sequence, e.g., antibiotic tolerance genes. The plasmid with the T-DNA and the expression cassette are then transfected into *Agrobacterium tumefaciens*. For effective transformation of plant cells, the *A.*

*tomographicus* bacterium also comprises the necessary *vir* regions on a native Ti plasmid. In an alternative transformation technique, both the Ti-DNA sequences as well as the *vir* sequences are on the same plasmid. For a discussion of *A. tumefaciens* gene transformation, see, Firoozabady & Kitchie, PLANT CELL, TISSUE AND ORGAN CULTURE: FUNDAMENTAL METHODS, Gamborg & Phillips (Eds.), Springer Lab Manual (1995).

For transformation of the plants of this invention in one aspect, explants are made of the tissues of desired plants, e.g., leaves. The explants are then incubated in a solution of *A. tumefaciens* at about  $0.8 \times 10^9$  to about  $1.0 \times 10^9$  cells/mL for a suitable time, typically several seconds. The explants are then grown for approximately 2 to 3 days on suitable medium.

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, *et al.*, PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE, pp. 124-176, Macmillan Publishing Company, New York, 1983; and Binding, REGENERATION OF PLANTS, PLANT PROTOPLASTS, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee, *et al.*, *Ann. Rev. of Plant Phys.* 38:467-486 (1987). See also, Payne, Gamborg, Atlas, Sigma-LSRCCC and Sigma-PCCS, *all supra*.

After transformation with *Agrobacterium*, the explants are then transferred to selection media. One of still will realize that the selection media depends on which selectable marker was co-transformed into the explants. After a suitable length of time, transformants will begin to form shoots. After the shoots are about 1 to 2 cm in length, the shoots should be transferred to a suitable root and shoot media. Selection pressure should be maintained once in the root and shoot media.

The transformants will develop roots in 1 to about 2 weeks and form plantlets. After the plantlets are from about 3 to about 5 cm in height, they should be placed in sterile soil in filter pots. Those of skill in the art will realize that different acclimation procedures should be used to obtain transformed plants of different species. In a preferred embodiment, cuttings, as well as somatic embryos of transformed plants, after developing a root and shoot, are transferred to medium for establishment of plantlets. For a description of selection and regeneration of transformed plants, *see*, Dodds & Roberts, EXPERIMENTS IN PLANT TISSUE CULTURE, 3RD ED., Cambridge University Press (1995).

The transgenic plants of this invention can be characterized either genotypically or phenotypically to determine the presence of the shuffled gene.

phenotypic trait. A phenotypic trait is a physical characteristic of a plant determined by the genetic material of the plant in concert with environmental factors. The presence of shuffled DNA sequences can be detected as described in the preceding sections on identification of an optimized shuffled nucleic acid, e.g., by PCR amplification of the genomic DNA of a transgenic plant and hybridization of the genomic DNA with specific labeled probes. The survival of plants on exposure to a selected stress where lipid production or type helps cope with the stress can also be used to monitor incorporation of a lipid synthetic shuffled gene into the plant.

Plants which are transduced with shuffled nucleic acids as taught herein to achieve desirable lipid production. Essentially any plant can acquire lipid production by the techniques herein. Some suitable plants for modified lipid biosynthesis include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Pennisetum*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Anthriscum*, *Heracleolus*, *Nemesis*, *Petalagonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browallia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, *Malus*, *Apium*, and *Datura*, including sugarcane, sugar beet, cotton, fruit trees, and legumes. Especially suitable are grass family crops such as maize, wheat, barley, oats, alfalfa, rice, millet, rye and the like as well as oil producing crops such as rapeseed, sunflower (and other composite family members). Industrially important legume crops such as soybeans are also especially suitable.

In addition to plants, microbes, fungi and animals can be transduced with the shuffled nucleic acids of the invention. In addition to the references noted throughout, one of skill can find guidance as to animal cell culture in *Freshney Culture of Animal Cells, a Manual of Basic Technique*, third edition Wiley-Liss, New York (1994)) and the references cited therein provides a general guide to the culture of cells. See also, Kuchler, *et al.* (1977) *Biochemical Methods in Cell Culture and Virology*, Kuchler, R.J., Dowden, Hutchinson and Ross, Inc., and Inaba, *et al.*, *J. Exp. Med.*, 176:1693-1702 (1992). Additional information on cell culture is found in Ausubel, Sambrook and Berger, *supra*. Cell culture media are described in Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL. Generally, one of skill is fully able to transduce cells from animals, plants, fungi,

bacteria and other cells using available techniques. Moreover, one of skill can transduce whole organisms with shuffled nucleic acids using available techniques.

#### EXAMPLES

The following examples are illustrative and not limiting. One of skill will recognize a variety of non-critical parameters which can be altered to achieve essentially similar results.

#### EXAMPLE: MODIFICATION OF THE ACTIVITY AND SUBSTRATE SPECIFICITY OF A FATTY ACID CHAIN-LENGTH DETERMINING ENZYME THROUGH DNA SHUFFLING

An acyl-ACP thioesterase hydrolyzes the thioester bond linking the acyl group and ACP, thus releasing a free fatty acid. When a gene encoding the 12:0-ACP thioesterase isolated from the Bay tree was introduced into Canola, the lipid composition of the canola oil was altered from mainly 18:1 to 12:0 (laureate, an oil mostly found in tropical plants such as coconuts). The primary substrate of the enzyme is 12:0-ACP; however, it also hydrolyzes 14:0-ACP at a much lower rate (~10%). The hydrolysis of 12:0-ACP by the enzyme was modified by replacing several amino acids, creating a 14:0-ACP thioesterase.

In luminescence bacteria, such as *Vibrio harveyi* and *Photobacterium phosphoreum*, the channeling of fatty acids into a fatty aldehyde substrate for the bioluminescence reaction is catalyzed by a multienzyme complex which channels fatty acids through LuxD (a 14:0-ACP thioesterase), LuxE (a synthetase) and LuxC (a reductase). The Lux operon has been isolated and can be transformed into other hosts to allow cells to emit light through bioluminescence. Improving the efficiency of the Lux system is performed by shuffling the whole operon. The improved mutants are identified by eye or by established sensitive instrumental procedures.

Fig. 4 provides a schematic of the *Vibrio Lux* system. The following table provides the response in photon yield of M17 cells to different fatty acids for the *Vibrio Lux* system.

Lipid	Relative Luminescence	Total quanta X 10 <sup>9</sup>
Ethanol	<0.0001	<1
Decanoic acid	0.3	100
Dodecanoic acid (12:0)	2.0	400

Tridecanoic acid (14:0)	56	6000
Tetradecanoic acid (14:0)	100	15000
Pentadecanoic acid	45	5000
Hexadecanoic acid	5.5	150
Octadecanoic Acid	0.1	<1
Myristoleic acid (14:1)	8.5	2000
Linoleic acid	0.01	<1

#### Recloning LuxD with shuffled plant 12:0-ACP thioesterase.

Because the wild-type thioesterase hydrolyzes 14:0-ACP in a low-rate, it is unlikely to be efficient to supply 14:0 for the Lux system. On the other hand, if a relatively efficient 14:0-ACP is generated by shuffling, it restores or improves the bioluminescence reaction. All genes involved in this example are published and obtainable (and, of course, any sequence can be used in *in silico* or oligonucleotide synthetic approaches as noted supra). See, e.g., Voelker TA et al. (1992) *Science* 257:72-74; Miyamoto CM et al. (1988) *J. Biol. Chem.* 263:13333-13399. Hosts for the lux system and measurement of the bioluminescence are also well documented.

#### 10 EXAMPLE: SHUFFLING FAS GENES TO INCREASE OIL YIELD OR TO INCREASE YIELD OF SPECIFIC FATTY ACIDS.

One possible limitation to the oil content of seeds is the rate of fatty acid synthesis. Shuffling the genes encoding enzymes in this pathway is a method of increasing oil yield. Further, the production of specific or unusual fatty acids can be limited by the ability of the FAS enzymes to utilize the specific fatty acids. In this case, shuffling or other diversity generation methods as noted herein can be used to increase the pathway flux and increase the yield of the specific fatty acids. The shuffled genes are expressed in oil bearing tissues using promoters that give preferential expression of the genes in those tissues (e.g. using a promoter that gives high expression in maturing embryos to drive expression of genes in oilseeds that produce oil in their embryos. This includes promoters such as phaseolin or napin to alter the oil content of oilseeds, such as soybean or canola.) The enzymes to be shuffled for this example can include, for example, acyl-ACP carboxylase, Keto acyl-ACP synthases, keto acyl-ACP reductases, hydroxyacyl-ACP dehydratases, and enoyl-ACP reductases.

#### EXAMPLE: SHUFFLING THE KENNEDY PATHWAY TO INCREASE OIL YIELD OR TO INCREASE YIELD OF SPECIFIC FATTY ACIDS.

The endoplasmic reticulum localized enzymes of the Kennedy pathway synthesize triglycerides from glycerol-3-phosphate and acyl-CoAs. Thus, shuffling the genes encoding enzymes in this pathway is a method to increase oil yield. Further, the production of specific or unusual fatty acids can be limited by the ability of the Kennedy pathway enzymes to utilize the specific fatty acids. In this case, shuffling and/or other diversity generation methods, as described herein, can be used to increase the pathway flux and increase the yield of the specific fatty acids. The shuffled genes are expressed in oil bearing tissues using promoters that give preferential expression of the genes in those tissues (e.g. using a promoter that gives high expression in maturing embryos to drive expression of genes in oilseeds that produce oil in their embryos. This includes promoters such as phaseolin or napin, to alter the oil content of oilseeds, such as soybean or canola.) The enzymes to be shuffled for this example include, e.g., glycerol-3-phosphate acyltransferase, lysophosphatidyl choline acyltransferase, phosphatidic acid phosphatase, diacylglycerol acyltransferase and the like.

#### EXAMPLE: SHUFFLING ACRYL-ACP DESATURASES TO YIELD NOVEL UNSATURATED FATTY ACIDS.

Acyl-ACP desaturases can introduce double bonds at different positions on 18 carbon fatty acids such as stearic acid, and can also introduce double bonds on other chain length fatty acids such as palmitic acid when the substrate fatty acids are esterified to ACP. There are a wide variety of acyl-ACP desaturases available as parents. Acyl-ACP desaturases that introduce double bonds at different positions on acyl chains or that use fatty acids of different chain lengths as substrates can be evolved using the techniques herein to provide evolved desaturases that can produce novel unsaturated fatty acids.

#### SHUFFLING MEMBRANE-ASSOCIATED DESATURASES TO YIELD NOVEL FATTY ACIDS

A family of fatty acid desaturases related to the desaturases that form linoleic and linolenic acids in plants can form novel fatty acids such as hydroxy fatty acids, acetylenic fatty acids, epoxy fatty acids, and fatty acids with conjugated cis and trans double bonds. Suitable parents can be identified using Arabidopsis FAD2 as a query to identify related sequences from Genbank™ (e.g., using BLAST or other

suitable search/alignment algorithms). The family can be evolved using the methods described herein to develop enzymes capable of forming novel fatty acids with the above structures at different places on fatty acid molecules, and on fatty acids of different chain lengths. Providing this class of enzymes can also result in the production of novel functional groups not known to exist on fatty acids in nature.

#### EXAMPLE: IMPROVING EFFICIENCY OF TRANSIT PEPTIDES GUIDING PROTEINS INTO CHLOROPLASTS OR ENDOPLASMIC RETICULUM

Most of the enzymes involved in fatty acid (FA) synthesis are imported into chloroplasts or ER. Most major systems that transport proteins across a membrane share the following features: an N-terminal transient signal sequence on the transported protein, a targeting system on the cis side of the membrane, a hetero-oligomeric transmembrane channel that is gated both across and within the plane of the membrane, a peripherally attached protein translocation motor that is powered by the hydrolysis of ATP, and a protein folding system on the trans side of the membrane.

Genetic engineering of FA synthesis commonly utilizes expression of genes from different species or even different kingdoms. In many cases, transit peptides of these enzymes do not efficiently guide the recombinant proteins into plastids of the new hosts, e.g. a transit peptide from Soybean may not work efficiently in *Canola*. In addition, some desired target enzymes (e.g. bacterial or some fungal ones) do not have signal peptides, and one is typically engineered for plant expression. Some transit peptides have been used successfully for recombinant expression, however, their effectiveness can not be generally applied. In some cases, large amounts of unprocessed (i.e., due to failure to import) proteins are detected, and the results have been speculated as a cause for poor phenotypes. It is desirable for transit peptides to be engineered to be highly specific and efficient.

Chloroplastic transit peptide sequences, which can vary in length from 20 to 120 amino acids, contain no obvious blocks of conserved amino acid sequence or secondary structure. In general, the N-proximal portion lacks both positively charged residues as well as glycine and proline. The central domain lacks acidic residues and is rich in hydroxylated amino acids such as serine and threonine. The C-terminal domain has a loosely conserved consensus sequence Ile/Val-X-Ala/Cys-Ala close to the cleavage site. Plastid, but not mitochondrial precursor proteins, are phosphorylated at the serine or threonine within the transit peptide by a cytosolic protein kinase.

In one embodiment, a group of transit peptides similar to that of the small subunit of ribulose-bisphosphate carboxylase is shuffled and cloned into the N-terminal domain of a reporter protein. The chimeric gene is cloned into an expression vector for expression in either *E. coli* or *cynobacteria*. The *cynobacteria* expression library is transformed into *Synechocystis*. Import is monitored by the expression of the reporter. The high performers are subjected to a second round import study in *Synechocystis* or by in vitro import experiments using isolated chloroplasts with *E. coli* produced chimeras.

#### EXAMPLE: MODIFICATION OF THE *PSEUDOMONAS PUTIDA CIS-TRANS ISOMERASE*, AN ENZYME INFLUENCING MEMBRANE FLUIDITY IN RESPONSE TO ORGANIC COMPOUNDS AND TEMPERATURE

Cells of *Pseudomonas putida* change the ratio of *cis* and *trans* monounsaturated fatty acids in response to growth temperature or membrane active compounds such as phenol or alcohol. These detoxification (or anti-stress) responses are attributed to a *cis-trans* isomerase. Different strains of *P. putida* have been isolated from various highly stressful environments. The specificity of the enzyme is relatively narrow. For example, the isomerase from *Pseudomonas* sp. Strain E-3 converts double bonds at positions 9, 10 or 11 but not 6 or 7 of *cis*-monounsaturated fatty acids with chain lengths of 14, 15, 16, and 17. However, 18:1 with double bonds at positions 9 or 11 are not substrates. On the other hand, the isomerase from strain P8 catalyzes the conversion at position 9 of 18:1. Furthermore, some *trans*-fatty acids are of commercial interest. Therefore a library of such *cis-trans* isomerase is made to allow selection activities in a number of applications.

#### Construction Of An Isomerase Library

A few *P. putida* isomerase genes have been cloned. PCR primers are designed to amplify large numbers of *P. putida* isomerases, and the amplified fragments are used for shuffling.

The resulting library is used for expression in *E. coli* or *Pseudomonas*. The recombinant organisms are screened under a various conditions, such as high concentration of organic compounds and elevated temperatures.

# EXAMPLE APPLICATION OF SHUFFLING IN ONE- AND TWO-HYBRID SYSTEMS

A one-hybrid assay is based on an interaction between a target-specific DNA-binding domain and a target-independent activation domain. The assay enables the rapid identification of novel DNA-binding proteins and access of their genes. The two-hybrid system enables the detection of protein-protein interaction and subsequent isolation of their genes.

In plant plastids, where lipid synthesis occurs, high levels of acyl-ACP are found. Certain genes encoding enzymes utilizing other acyl- molecules (e.g. acyl-CoA), therefore, are incomparable for catalysis in plastids. For example, it is desirable for PHB polymerase to utilize PHB-ACP other than its natural substrate PHB-CoA to achieve high level production of the biopolymer in plants.

In one embodiment, the substrate specificities of enzymes are modified from utilizing CoA or other linker molecules to acyl-carrier protein (ACP). These enzymes include, but not are limited to, desaturase, isomerase, thioesterase, and PHA polymerase. It is also desirable to modify regulatory elements, transcription factors and signal transduction elements for improved specificity and cross-species recognition. Examples of one and two hybrid shuffling protocols relevant to the present invention are found in Figures 3A and B.

As shown in Figure 3A, a two-hybrid system can be used for screening. For example, KAS proteins, which are known to form heterodimers, resulting in varied substrate specificities can be evolved. Similarly, PHA polymerase can be modified to use PHA-ACP instead of PHA-CoA. Other enzymes (PKS, Desaturases, TE, ACCase, etc.) can also be similarly screened in a two-hybrid system. As shown, genes for a target protein (X) are cloned into a DNA-BD vector. A random collection of cDNA (Y) is cloned into an AD vector. Both plasmids are co-transformed into yeast. Hybrid proteins are expressed in the same cell.  $\beta$ -galactosidase activity is screened for to confirm interacting proteins.

As shown in Fig. 3B, a one-hybrid system is used to acquire modified genes of interest. As shown, shuffled transcription factors (TF-AD) are screened on a known target element ("E") connected to a known reporter system. For example, the Napin promoter, which is a strong seed-specific promoter can be used (for some targets, the Napin promoter is activated too late or too early, depending on the target). A

transcription factor controlling an earlier promoter can be modified to bind the napin element. The napin transcription factor can be modified to bind a weaker promoter.

## Example: Zinc Finger Libraries

Genomes are regulated at the level of transcription, primarily through the action of transcription factors that bind DNA in a sequence-specific fashion.

Transcription factors frequently act both through a DNA-binding domain that localizes the protein to a specific site within the genome and through accessory effector domains that act to activate or repress transcription at or near that site. Zinc finger proteins, a Cys2-His2 class of nucleic acid-binding proteins, have unique structural features that allow recognition of specific DNA sequences, and can be engineered as fusion proteins with either an activator domain or a repressor domain (Beetli RR et al. 1998. Proc. Natl. Acad. Sci. USA. 95:14628). The artificial transcriptional regulators can repress or activate gene expression in a specific manner. Results also indicated that gene activation or repression was achieved by targeting within the gene transcript, suggesting that information obtained from expressed sequence tags (ESTs) is sufficient for the construction of gene switches. These switches are useful, e.g., for controlling lipid synthetic enzymes.

Construction of Zinc Finger Libraries.

Poly(dactyl) zinc finger proteins are constructed from modular building blocks (Beetli et al. 1998, *id.*). These building blocks are substrates for DNA shuffling. Libraries are made by shuffling individual blocks or shuffling blocks in combination, generating greater diversity of zinc finger proteins than simple genomic PCR based assembly methods. The shuffled zinc finger DNAs are cloned into vectors with or without sequences encoding either activators or repressors.

## Identification of zinc fingers recognizing a specific DNA sequence.

A specific DNA recognition sequence can be, e.g., in the 5' untranslated region or 5' translated region of a known gene such as a lipid synthetic enzyme gene. It also can be a known promoter sequence, or an EST of interest as well as an EST with no homology found with other genes. This specific sequence is cloned as a fusion with a reporter gene under the control of an appropriate promoter. The hosts containing this vector are transformed with a library of shuffled zinc finger DNAs. The expression of the reporter gene is monitored. A particular zinc finger protein is identified by altered reporter gene expression due to the binding of the specific sequence by zinc finger. In

other cases, a specific DNA sequence can be cloned without fusing with a reporter gene, the presence or absence of the transcripts of this DNA may be detected by hybridization.

Identification of zinc fingers associated with a particular phenotype.  
Library of zinc fingers are transformed into prokaryotes (*E. coli*, etc.), or eukaryotes (fungi, plants, animals, etc.) and selected for desirable phenotypes. The production of the zinc fingers is controlled by appropriate promoters of desired timing and specificity. The zinc fingers may be fused with either an activator or a repressor.

Identification of zinc finger DNA-binding specificities by ELISA.  
Cell lysates containing recombinant zinc finger proteins are used in ELISA assays with immobilized biotinylated hairpin oligonucleotides containing specific sequences. A high-throughput system is used for this process. Identification of zinc fingers can also be performed using a one-hybrid system as above.

## 15 EVOLUTION OF METHYLTRANSFERASES TO PRODUCE BRANCHED CHAIN FATTY ACIDS, METHOXY FATTY ACIDS AND OTHER UNUSUAL FATTY ACIDS.

This example provides methods to evolve methyltransferases, particularly cyclopropane fatty acid synthase related enzymes, to form branched chain fatty acids. Branched chain fatty acids have the physical characteristics of unsaturated fatty acids, yet they have the oxidative stability of saturated fatty acids. Thus, they have desirable properties as industrial oils, and they may have some food oil applications. According to the present invention, shuffled genes, e.g., derived from bacterial cyclopropane fatty acid synthases, can be used to transform oilseed crops to produce oils with branched chain fatty acids. Additionally, novel enzymes that can form cyclopropyl fatty acids, methoxy fatty acids or keto fatty acids can also be made by this approach.

Biochemical characterization of 10-methyl stearate biosynthesis in *Mycobacterium* suggests that this branched chain fatty acid is derived from addition of a methyl group from SAM to oleic acid esterified to phospholipids (Akamatsu and Law (1968) *Biochem. Biophys. Res. Commun.* 33:172-176; Akamatsu and Law *Biol. Chem.* 245:701-708). More recent molecular biology has shown that a class of SAM methyltransferases add methyl groups to double bonds of fatty acids to form cyclopropane fatty acids (CFA) in *E. coli* (reviewed in Grogan and Cronan (1997)

*Microbiology and Molecular Biology Reviews* 61:429-441) and *Mycobacterium* (Yuan et al. (1998) *J. Biol. Chem.* 273:21282-21290; Barry et al. (1998) *Proc. Lipid Res.* 37:143-79). A group of related genes from *Mycobacterium* also form methoxy fatty acids, methylene branched fatty acids, methyl branched fatty acids, and keto fatty acids. Some forms of the enzymes act on fatty acids esterified to ACP or to fatty acids esterified to glycerol in phospholipids. These enzymes all act by the addition of methyl groups to double bonds of unsaturated fatty acids. The enzyme involved in the formation of the branched chain fatty acid, 10-methyl stearate, has not been identified. It is unknown whether this is a minor side activity of one of the known CFA type enzymes, or if an unknown enzyme is dedicated to the synthesis. In addition to the genes from *E. coli* and *Mycobacterium*, whose protein products have been characterized biochemically, there are numerous accessions of related genes from multiple bacteria deposited in Genbank.

There is a large family of putative methyl transferases related to CFA synthases deposited in Genbank. The parent sequences are therefore, easy to obtain and use in the shuffling procedures of the invention. Parent sequences are identified, e.g., by using the *E. coli* CFA synthase as a query against Genbank or other protein or nucleotide databases. A number of CFA synthase related parents are isolated, and used for DNA shuffling or the other diversity generation procedures noted herein. The shuffled library is cloned into an *E. coli* expression vector. The library is transformed, e.g., into an *E. coli* mutant deficient in the synthesis of unsaturated fatty acids (*fabB*). This strain requires supplementation of unsaturated fatty acids in the growth medium, and thus can be fed oleic acid. Oleic acid containing phospholipids would be suitable substrates for evolved methyltransferases, and provide a suitable screening system to predict the phenotype one observes in transgenic plant oils. The shuffled library is screened, e.g., using gas chromatography, to detect branched chain fatty acids and other unusual fatty acids such as keto or methoxy fatty acids.

Enzymes with desired methyltransferase activities identified through screening in *E. coli* are then tested for their ability to modify plant oils by expression in transgenic *Arabidopsis* plants.

Parents useful as substrates in shuffling or other diversity generation reactions are selected, e.g., from the following list of CFA synthase related genes (accession numbers are indicated):

30	glt3144826 (AET004296) cyclopropane fatty acid synthase [Pho...]	804	0.0
	glt30001063CA_ECOLI CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SY...	804	0.0
	glt290474821 (AEB000357) cyclopropane fatty acid synthase ...	206	36-55
	glt2954038 (AED00753) cyclopropane-fatty-acyl-phospholipid ...	206	36-55
	glt4193552 (AEB01326) CYCLOPROPANOCYCLOPROPANE-FATTY ACID SYNTHA...	206	36-55
5	gltAAAE11721.LJAE002051_7 (AED002051) cyclopropane fatty-acyl...	180	26-6
	glt310499723_PSEBU HYDROLYTICAL 44.7 KD PROTEIN IN LFD-...	166	36-4
	glt3CAAI17269 (ALD019932) uha1 [Mycobacterium tuberculosis]	154	16-33
10	glt3CAAI19156A1 (ALD023596) hydrophobic protein MLCB2407.16...	153	26-33
	glt3CAB07103 (Z92772) mmuA4 [Mycobacterium tuberculosis]	153	26-33
	glt1825533 (U77466) CmaB [Mycobacterium bovis BCG]	153	26-33
	glt1575542 (U66108) mycolic acid synthase 2 [Mycobac...	153	26-33
	glt111966CEA2 MYCYU CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID S...	151	96-33
	glt1006709 (U34637) cyclopropane mycolic acid synthase 2 [M...	151	96-33
15	glt3ST72836_E2168_F3_130 protein - Mycobacterium lepre - Pg...	150	36-3
	glt3CACA18092A1 (ALD022121) hydrophobic protein Rv3170 [Myc...	150	36-3
	glt3CACA2213411 (ALD013355) cDNA EST yk2011.5 comes from th...	144	66-33
	glt111951CEFA1 MYCYU CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID S...	144	16-33
	glt4355091CEFA CITER CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SY...	143	26-33
20	glt3CACA1174251 (ALD021933) mmuA4 [Mycobacterium tuberculosis]	140	26-33
	glt1575548 (U66108) mycolic mycolic acid synthase 3 [Mycoba...	140	26-33
	glt3CACA1174251 (ALD021933) mmuA1 [Mycobacterium tuberculosis]	138	16-33
	glt397878248 (AAT01078) mycolic acid methyl transferase [Myc...	136	4-33
	glt1825532 (U77466) CmaB [Mycobacterium bovis BCG] -glt31261...	131	86-33
25	glt1575546 (U66108) mycolic mycolic acid synthase 3 [Mycoba...	130	26-33
	glt3CAB071011 (Z927772) mmuA4 [Mycobacterium tuberculosis]	129	46-33
	glt1575549 (U66108) mycolic mycolic acid synthase 4 [Mycoba...	128	76-33
	glt1825532 (U77466) CmaB [Mycobacterium bovis BCG]	128	96-33
	glt1825353 (U77466) CmaB [Mycobacterium bovis BCG]	121	16-2
	glt3CAB367826A1 (ALD035525) putative protein [Arabisopsis th...	112	66-6
	glt3CAB367835A1 (ALD035525) putative protein [Arabisopsis th...	112	36-1
	glt2114826 (AET004296) cyclopropane fatty acid synthase [Pho...	67	26-1

herein; (3) one or more lipid assay component; (4) a container for holding lipids, nucleic acids, plants, cells, or the like and, (5) packaging materials.

In a further aspect, the present invention provides for the use of any component or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit to practice any assay or method herein.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and materials described above can be used in various combinations. All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.

WHAT IS CLAIMED IS

1. A method of making a nucleic acid encoding a lipid biosynthetic activity, the method comprising:  
recombining a plurality of parental nucleic acids to produce one or more recombinant lipid biosynthetic nucleic acids comprising a distinct or improved lipid biosynthetic activity; and,  
selecting the one or more recombinant lipid biosynthetic nucleic acid for one or more encoded lipid biosynthetic activity or,  
selecting the one or more recombinant lipid biosynthetic nucleic acid for enhanced or reduced encoded polypeptide expression or stability;  
thereby producing a selected shuffled lipid biosynthetic nucleic acid, which nucleic acid encodes a selected lipid biosynthetic activity.
2. The method of claim 1, wherein the lipid biosynthetic activity is selected from:  
modulation of lipid saturation for one or more selected lipids produced by a lipid synthetic pathway comprising activity encoded by the one or more selected shuffled lipid biosynthetic nucleic acids,  
modulation of fatty acid composition in a transgenic plant, algae, animal, bacterium or fungus expressing the selected shuffled lipid biosynthetic nucleic acid,  
modulation of fatty alcohol composition in a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid,  
modulation of a wax composition in a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid,  
modification of acyl chain length in a lipid produced by a lipid synthetic pathway comprising activity encoded by the selected shuffled lipid biosynthetic nucleic acid,  
location of fatty acid accumulation in a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid,  
modulation of triglyceride yield of a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid.

- modulating the activity, expression or localization of one or more endoplasmic reticulum localized enzyme of the Kennedy pathway,  
an increased ability of a molecule encoded by the selected shuffled lipid biosynthetic nucleic acid, or a cell transduced with the selected shuffled lipid biosynthetic nucleic acid to chemically modify a lipid or lipid precursor,  
an increase or alteration in the range of lipid substrates for a cell transduced with the selected shuffled lipid biosynthetic nucleic acid,  
an increased expression level of a lipid biosynthetic polypeptide in a cell transduced with the selected shuffled lipid biosynthetic nucleic acid,  
a decrease in susceptibility of a lipid biosynthetic polypeptide in a cell transduced with the selected shuffled lipid biosynthetic nucleic acid to protease cleavage,  
a decrease in susceptibility of a lipid biosynthetic polypeptide encoded by the selected shuffled lipid biosynthetic nucleic acid in a cell to high or low pH levels,  
a decrease in susceptibility of a protein encoded by the selected shuffled lipid biosynthetic nucleic acid in a cell to high or low temperatures,  
a decrease in toxicity to a cell by a lipid biosynthetic polypeptide encoded by the selected shuffled lipid biosynthetic nucleic acid, as compared to one of the parental nucleic acids, when expressed in a cell, and,  
a modification in the activity of a methyltransferase, which modification results in the formation of branched chain fatty acids, cyclopropyl fatty acids, methoxy fatty acids, or keto fatty acids by the methyltransferase.
3. The method of claim 1, wherein the activity of the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid is selected by detecting one or more of a change in a physical property of one or more lipid, fatty acid, wax or oil in the presence of a polypeptide or RNA encoded by the one or more lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, a protein-protein interaction in a two hybrid assay, expression of a reporter gene in a one hybrid assay, growth or survival of a recombinant cell expressing the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in an elevated temperature environment, growth or survival of a recombinant cell expressing the one or more recombinant lipid

biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a medium comprising a membrane active compound, relative bioluminescence of a recombinant cell comprising at least one gene from the Lux operon and the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid to a chloroplast, or endoplasmic reticulum, and detection of cellular localization of a product produced as a result of expression of the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a cell.

4. The method of claim 1, wherein at least one of the parental nucleic acids is the same as, or homologous to, a nucleic acid encoding a protein selected from: an Acetyl-CoA carboxylase (an ACCase), a homomeric acetyl-CoA carboxylase, a heteromeric acetyl-CoA carboxylase BC subunit, a heteromeric acetyl-CoA carboxylase, a BCCP subunit, a heteromeric acetyl-CoA carboxylase (alpha)-CT subunit, a heteromeric acetyl-CoA carboxylase (beta)-CT subunit, an acyl carrier protein (ACP) (plastidial isoform or mitochondrial isoform), a malonyl-CoA:ACP transacylase, a ketoacyl-ACP synthase (KAS), a KAS I, a KAS II, a KAS III, a ketoacyl-ACP reductase, a 3-hydroxyacyl-ACP, an enoyl-ACP reductase, a stearyl-ACP desaturase, an acyl-ACP thioesterase (Fad), a FadA, a FadB, a glycerol-3-phosphate acyltransferase, a 1-acyl-sn-glycerol-3-phosphate acyltransferase, a plastidial cytidine-5'-diphosphate-diacylglycerol synthase, a plastidial phosphatidylglycerol-phosphate synthase, a plastidial phosphatidylglycerol-3-phosphate phosphatase, a phosphatidylglycerol desaturase (palmitylate specific), a plastidial oleate desaturase (fad6), a plastidial linoleate desaturase (fad7/fad8), a plastidial phosphatidic acid phosphatase, a monogalactosyldiacylglycerol synthase, a monogalactosyldiacylglycerol desaturase (palmitylate-specific), a digalactosyldiacylglycerol synthase, a sulfolipid biosynthesis protein, a long-chain acyl-CoA synthetase, an ER glycerol-3-phosphate acyltransferase, an ER 1-acyl-sn-glycerol-3-phosphate acyltransferase, an ER phosphatidic acid phosphatase, a diacylglycerol cholinephosphotransferase, an ER

oleate desaturase (fad2), an ER linoleate desaturase (fad3), an ER cytidine-5'-diphosphate-diacylglycerol synthase, an ER phosphatidylglycerol-phosphate synthase, an ER phosphatidylglycerol-3-phosphate phosphatase, a phosphatidylinositol synthase, a diacylglycerol kinase, a cholinephosphate cytidylyltransferase, a phosphatidylcholine transfer protein, a choline kinase, a Lipase, a phospholipase C, a phospholipase D, a phosphatidylserine decarboxylase, a phosphatidylinositol-3-kinase, a ketoacyl-CoA synthase (KCS), a (beta)-keto-acyl reductase, a CER2, a fatty acid isomerase an enzyme which forms a cyclopropyl fatty acid, an enzyme which forms a methoxy fatty acid, an enzyme which forms a keto fatty acid, and a bacterial cyclopropane fatty acid synthase.

5. The method of claim 1, wherein at least one of the parental nucleic acids is the same as, or homologous to, a nucleic acid encoding a protein which affects oil yield, which protein is selected from: an ACCase, an sn-2 acyltransferase, an acyltransferase other than sn-2 acyltransferase, a malonyl-CoA:ACP transacylase, an oleosin, a fatty acid binding protein, an Acyl-CoA synthase, and an acyl-ACP synthase.

6. The method of claim 1, wherein at least one of the parental nucleic acids is the same as, or homologous to, a nucleic acid encoding a protein which affects fatty acid acyl chain length or composition, which protein is selected from: a thioesterase and an elongase.

7. The method of claim 1, wherein at least one of the parental nucleic acids is the same as, or homologous to, a nucleic acid encoding a protein which affects fatty acid saturation, which protein is selected from: a desaturase, a cis-trans isomerase, and a lipoygenase.

8. The method of claim 1, wherein at least one of the parental nucleic acids is the same as, or homologous to, a nucleic acid encoding a protein which affects fatty acid branch structures, which protein is a reductase, or a lipoygenase.

9. The method of claim 1, wherein at least one of the parental nucleic acids is the same as, or homologous to, a nucleic acid encoding a protein which affects flavor, which protein is selected from: a Lox protein, a desaturase, a beta-oxidation enzyme, and a hydroperoxide lyase.

10. The method of claim 1, wherein at least one of the parental nucleic acids is the same as, or homologous to, a nucleic acid encoding a protein which affects polyunsaturation, which protein is selected from: a protein in the PKS-like operon, a desaturase, an elongase, a lipase and a DNA binding protein.

11. The method of claim 1, wherein the parental nucleic acids are homologous.

12. The method of claim 1, wherein the parental nucleic acids are at least about 60% identical.

13. The method of claim 1, wherein at least one of the parental nucleic acids does not encode a lipid biosynthetic activity.

14. The method of claim 1, wherein any of the parental nucleic acids, the one or more the one or more recombinant lipid biosynthetic nucleic acid, and the selected recombinant lipid biosynthetic nucleic acid, is cloned into an expression vector.

15. The method of claim 1, wherein the plurality of parental nucleic acids are shuffled to produce a library of recombinant nucleic acids comprising one or more library member nucleic acid encoding one or more lipid biosynthetic activity, which library is selected for one or more lipid biosynthetic activity selected from: a change in a physical property of one or more lipid, fatty acid, wax or oil in the presence of a polypeptide or RNA encoded by one or more library member, a protein-protein interaction in a two hybrid assay comprising a library member, expression of a reporter gene in a one hybrid assay performed on a library member, growth or survival of a recombinant cell expressing one or more library member, growth or survival of a recombinant cell expressing one or more library member in a medium comprising a membrane active compound, relative bioluminescence of a recombinant cell comprising a *lux* operon gene member and one or more library member, detection of cellular localization of a protein encoded by one or more library member, and detection of cellular localization of a product produced as a result of expression of one or more recombinant lipid biosynthetic nucleic acid in a cell.

16. The method of claim 1, wherein a plurality of recombinant lipid synthetic nucleic acids are produced, thereby providing a library of recombinant nucleic acids comprising one or more lipid biosynthetic activity.

17. The library made by the method of claim 16, wherein the library is a phage display library.

18. The library made by the method of claim 16, wherein the library is present in one or more cell selected from an *E. coli*, a cyanobacteria and a *Synechocystis*.

19. The library made by the method of claim 18, wherein the library is first made in an *E. coli* or a cyanobacteria and then transduced into a *Synechocystis*.

20. The library made by the method of claim 16, wherein the library is present in *Pseudomonas putida* cells.

21. The library made by the method of claim 16, wherein the parental nucleic acids are shuffled in a plurality of cells, which cells are prokaryotes or eukaryotes.

22. The library made by the method of claim 16, wherein the parental nucleic acids are shuffled in a plurality of cells, which cells are plants, yeast, bacteria, or fungi.

23. The method of claim 1, wherein the parental nucleic acids are shuffled in a plurality of cells; the method optionally further comprising one or more of:

(a) recombining DNA from the plurality of cells that display lipid biosynthetic activity with a library of DNA fragments, at least one of which undergoes recombination with a segment in a cellular DNA present in the cells to produce recombined cells, or recombining DNA between the plurality of cells that display lipid biosynthetic activity to produce cells with modified lipid biosynthetic activity;

(b) recombining and screening the recombined or modified cells to produce further recombined cells that have evolved additionally modified lipid biosynthetic activity; and,

(c) repeating (a) or (b) until the further recombined cells have acquired a desired lipid biosynthetic activity.

24. The method of claim 1, wherein the method further comprises:

(a) recombining at least one selected shuffled lipid biosynthetic nucleic acid with a further lipid biosynthetic activity nucleic acid, which further nucleic acid is the same or different from one or more of the plurality of parental nucleic acids, thereby producing a library of recombinant lipid biosynthetic nucleic acids;

(b) screening the library to identify at least one further selected distinct or improved recombinant lipid biosynthetic nucleic acid that exhibits a further improvement or distinct property compared to the plurality of parental nucleic acids; and, optionally,

(c) repeating (a) and (b) until the resulting additional further distinct or improved recombinant nucleic acid shows an additionally distinct or improved lipid biosynthetic property.

25. The method of claim 1, wherein the one or more recombinant lipid biosynthetic nucleic acid is present in one or more bacterial, yeast, plant or fungal cells and the method comprises:

pooling multiple separate lipid biosynthetic nucleic acids; screening the resulting pooled lipid biosynthetic nucleic acids to identify distinct or improved recombinant lipid biosynthetic nucleic acids that exhibit distinct or improved lipid biosynthetic activity compared to a non-recombinant lipid biosynthetic activity nucleic acid; and, cloning the distinct or improved recombinant nucleic acid.

26. The method of claim 25, further comprising transducing the resulting cloned distinct or improved nucleic acid into a prokaryotic or eukaryotic.

27. The method of claim 1, wherein recombining the plurality of parental nucleic acids is performed by family nucleic acid shuffling.

28. The method of claim 1, wherein recombining the plurality of parental nucleic acids comprises individual nucleic acid shuffling.

29. The method of claim 1, wherein recombining the plurality of parental nucleic acids comprises oligonucleotide-mediated nucleic acid shuffling.

30. The method of claim 1, wherein recombining the plurality of parental nucleic acids comprises *in silico* nucleic acid shuffling.

31. A selected shuffled lipid biosynthetic nucleic acid made by the method of claim 1.

32. A plant, bacteria or fungus transduced with the selected shuffled lipid biosynthetic nucleic acid of claim 1.

33. The plant of claim 32, wherein the plant is selected from the families Gramineae, Composite, and Leguminosae.

34. The plant of claim 32, wherein the plant is selected from corn, peanut, barley, millet, rice, soybean, sorghum, wheat, oats, sunflower, and a nut plant.

35. The plant of claim 32, wherein the plant exhibits a new lipid biosynthetic activity as compared to a wild-type non-transduced plant.

36. A DNA shuffling mixture, comprising: at least three homologous DNAs, each of which is derived from a nucleic acid encoding a polypeptide or polypeptide fragment which encodes a lipid biosynthetic activity.

37. The DNA shuffling mixture of claim 36, wherein the at least three homologous DNAs are present in cell culture or *in vitro*.

38. A method of modulating lipid biosynthetic activity in a cell, comprising: performing whole genome shuffling of a plurality of genomic nucleic acids in the cell and selecting for one or more lipid biosynthetic activity.

39. The method of claim 38, wherein the genomic nucleic acids are from a species or strain different from the cell.

40. The method of claim 38, wherein the cell is of prokaryotic or eukaryotic origin.

41. The method of claim 38, wherein the lipid biosynthetic activity to be selected is selected from: modulation of lipid saturation for one or more selected lipids produced by a lipid biosynthetic pathway comprising activity encoded by the one or more selected shuffled lipid biosynthetic nucleic acids, modulation of fatty acid composition in a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid, modulation of fatty alcohol composition in a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid, modulation of a wax composition in a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid, modification of acyl chain length in a lipid produced by a lipid biosynthetic pathway comprising activity encoded by the selected shuffled lipid biosynthetic nucleic acid, location of fatty acid accumulation in a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid, modulation of triglyceride yield of a transgenic plant, algae, animal, bacteria or fungus expressing the selected shuffled lipid biosynthetic nucleic acid, an increased ability of a molecule encoded by the selected shuffled lipid biosynthetic nucleic acid, or a cell transduced with the selected shuffled lipid biosynthetic nucleic acid, to chemically modify a lipid or lipid precursor, an increase or alteration in the range of lipid substrates for a cell transduced with the selected shuffled lipid biosynthetic nucleic acid, an increased expression level of a lipid biosynthetic polypeptide in a cell transduced with the selected shuffled lipid biosynthetic nucleic acid, a decrease in susceptibility of a lipid biosynthetic polypeptide in a cell transduced with the selected shuffled lipid biosynthetic nucleic acid to protease cleavage, a decrease in susceptibility of a lipid biosynthetic polypeptide encoded by the selected shuffled lipid biosynthetic nucleic acid in a cell to high or low pH levels, a decrease in susceptibility of a protein encoded by the selected shuffled lipid biosynthetic nucleic acid in a cell to high or low temperatures, and a decrease in toxicity to a cell by a lipid biosynthetic polypeptide encoded by the selected shuffled lipid biosynthetic nucleic acid, as compared to one of the parental nucleic acids, when expressed in a cell.

42. A method of obtaining a recombinant lipid biosynthetic nucleic acid which can confer modified lipid production to a plant in which the recombinant lipid biosynthetic nucleic acid is present, the method comprising:

73

(i) recombining a plurality of forms of a selected lipid synthetic nucleic acid which comprises segments derived from one or more parental nucleic acid which encode a lipid biosynthetic activity, or which can be shuffled to confer a lipid biosynthetic activity, wherein the plurality of forms of the selected nucleic acid differ from each other in at least one nucleotide, to produce a library of recombinant lipid biosynthetic nucleic acids; and,

(ii) screening the library to identify at least one recombinant lipid biosynthetic nucleic acid that exhibits distinct or improved lipid biosynthetic activity as compared to the parental nucleic acid.

43. The method of claim 42, wherein one or more parental nucleic acid encodes a lipid biosynthetic enzyme.

44. The method of claim 42, wherein the parental nucleic acids do not encode a lipid biosynthetic enzyme, wherein recombining the plurality of forms of a selected nucleic acid provides a nucleic acid which encodes a lipid biosynthetic enzyme.

45. The method of claim 42, wherein the plant is a crop.

46. The method of claim 42, wherein the selected nucleic acid encodes a polypeptide or polypeptide fragment selected from: an Acetyl-CoA carboxylase (an ACCase), a homomeric acetyl-CoA carboxylase, a heteromeric acetyl-CoA carboxylase BC subunit, a heteromeric acetyl-CoA carboxylase, a BCCP subunit, a heteromeric acetyl-CoA carboxylase (alpha)-CT subunit, a heteromeric acetyl-CoA carboxylase (beta)-CT subunit, an acyl carrier protein (ACP) (plastidial isoform or mitochondrial isoform), a malonyl-CoA:ACP transacylase, a ketoacyl-ACP synthase (KAS), a KAS I, a KAS II, a KAS III, a ketoacyl-ACP reductase, a 3-hydroxyacyl-ACP, an enoyl-ACP reductase, a stearyl-ACP desaturase, an acyl-ACP thioesterase (Fad), a FadA, a FadB, a glycerol-3-phosphate acyltransferase, a 1-acyl-sn-glycerol-3-phosphate acyltransferase, a plastidial cytidine-5'-diphosphate-diacylglycerol synthase, a plastidial phosphatidylglycero-phosphate synthase, a plastidial phosphatidylglycerol-3-phosphate phosphatase, a phosphatidylglycerol desaturase (palmate specific), a plastidial oleate desaturase (Fad6), a plastidial linoleate desaturase (Fad7/Fad8), a plastidial phosphatidic acid phosphatase, a monogalactosyldiacyl-glycerol synthase, a monogalactosyldiacyl-

74

- glycerol desaturase (palmitate-specific), a digalactosyl-diacyl-glycerol synthase, a sulfolipid biosynthesis protein, a long-chain acyl-CoA synthetase, an ER glycerol-3-phosphate acyltransferase, an ER 1-acyl-sn-glycerol-3-phosphate acyltransferase, an ER phosphatidic acid phosphatase, a diacylglycerol cholinephosphotransferase, an ER oleate desaturase (fad2), an ER linoleate desaturase (fad3), an ER cytidine-5'-diphosphate-diacylglycerol synthase, an ER phosphatidylglycerol-phosphate synthase, an ER phosphatidylglycerol-3-phosphate phosphatase, a Phosphatidylinositol synthase, a diacylglycerol kinase, a cholinephosphate cytidylyltransferase, a phosphatidylcholine transfer protein, a choline kinase, a Lipase, a phospholipase C, a phospholipase D, a phosphatidylserine decarboxylase, a phosphatidylinositol-3-kinase, a ketoacyl-CoA synthase (KCS), a (beta)-keto-acyl reductase, and a CER2, a fatty acid isomerase a fatty acid hydroxylase, a fatty acid epoxidase, a fatty acid acetylase, a methyl transferase related enzyme which alters lipid, a cyclopropane fatty acid synthase, a meromycolic acid synthase, a cyclopropane mycolic acid synthase, a diacylglycerol acyltransferase (DGAT), an acyl Co-A reductase, a wax synthase, a Cholesterol:Acyl-CoA acyltransferases (ACAT), and a lecithin:Acyl-CoA Acyltransferase (LCAT).

47. The method of claim 42, wherein the selected nucleic acid is derived from a parental nucleic acid derived from one or more gene encoding a protein selected from an Acetyl-CoA carboxylase (an ACCase), a homomeric acetyl-CoA carboxylase, a heteromeric acetyl-CoA carboxylase BC subunit, a heteromeric acetyl-CoA carboxylase (alpha)-CT subunit, a heteromeric acetyl-CoA carboxylase (beta)-CT subunit, an acyl carrier protein (ACP) (plastidial isoform or mitochondrial isoform), a malonyl-CoA:ACP transacylase, a ketoacyl-ACP synthase (KAS), a KAS I, a KAS II, a KAS III, a ketoacyl-ACP reductase, a 3-hydroxyacyl-ACP, an enoyl-ACP reductase, a stearyl-ACP desaturase, an acyl-ACP thioesterase (Fad), a FadA, a FadB, a glycerol-3-phosphate acyltransferase, a 1-acyl-sn-glycerol-3-phosphate acyltransferase, a plastidial cytidine-5'-diphosphate-diacylglycerol synthase, a plastidial phosphatidylglycerol-phosphate synthase, a plastidial phosphatidylglycerol-3-phosphate phosphatase, a phosphatidylglycerol desaturase (palmitate specific), a plastidial oleate desaturase (fad6), a plastidial linoleate desaturase (fad7/fad8), a plastidial phosphatidic acid phosphatase, a monogalactosyl-diacyl-glycerol synthase, a monogalactosyl-diacyl-

- glycerol desaturase (palmitate-specific), a digalactosyl-diacyl-glycerol synthase, a sulfolipid biosynthesis protein, a long-chain acyl-CoA synthetase, an ER glycerol-3-phosphate acyltransferase, an ER 1-acyl-sn-glycerol-3-phosphate acyltransferase, an ER phosphatidic acid phosphatase, a diacylglycerol cholinephosphotransferase, an ER oleate desaturase (fad2), an ER linoleate desaturase (fad3), an ER cytidine-5'-diphosphate-diacylglycerol synthase, an ER phosphatidylglycerol-phosphate synthase, an ER phosphatidylglycerol-3-phosphate phosphatase, a Phosphatidylinositol synthase, a diacylglycerol kinase, a cholinephosphate cytidylyltransferase, a phosphatidylcholine transfer protein, a choline kinase, a Lipase, a phospholipase C, a phospholipase D, a phosphatidylserine decarboxylase, a phosphatidylinositol-3-kinase, a ketoacyl-CoA synthase (KCS), a (beta)-keto-acyl reductase, and a CER2, a fatty acid isomerase a fatty acid hydroxylase, a fatty acid epoxidase, a fatty acid acetylase, a methyl transferase related enzyme which alters lipid, a cyclopropane fatty acid synthase, a meromycolic acid synthase, a cyclopropane mycolic acid synthase, a diacylglycerol acyltransferase (DGAT), an acyl Co-A reductase, a wax synthase, a Cholesterol:Acyl-CoA acyltransferases (ACAT), and a lecithin:Acyl-CoA Acyltransferase (LCAT).

48. The method of claim 42, wherein the plurality of forms of the selected nucleic acid comprise allelic or interspecific variants of the selected nucleic acid.
49. The method of claim 42, wherein the plurality of forms of the selected nucleic acids is produced by synthesizing or expressing a plurality of homologous nucleic acids.

50. The method of claim 42, wherein the library comprises a nucleic acid homologous to a nucleic acid encoding or partially encoding: an Acetyl-CoA carboxylase (an ACCase), a homomeric acetyl-CoA carboxylase, a heteromeric acetyl-CoA carboxylase BC subunit, a heteromeric acetyl-CoA carboxylase, a BCCP subunit, a heteromeric acetyl-CoA carboxylase (alpha)-CT subunit, a heteromeric acetyl-CoA carboxylase (beta)-CT subunit, an acyl carrier protein (ACP) (plastidial isoform or mitochondrial isoform), a malonyl-CoA:ACP transacylase, a ketoacyl-ACP synthase (KAS), a KAS I, a KAS II, a KAS III, a ketoacyl-ACP reductase, a 3-hydroxyacyl-ACP, an enoyl-ACP reductase, a stearyl-ACP desaturase, an acyl-ACP thioesterase

(Fat), a FatA, a FatB, a glycerol-3-phosphate acyltransferase, a 1-acyl-sn-glycerol-3-phosphate acyltransferase, a plastidial cytidine-5'-diphosphate-diacylglycerol synthase, a plastidial phosphatidylglycero-phosphate synthase, a plastidial phosphatidylglycerol-3-phosphate phosphatase, a phosphatidylglycerol desaturase (palmite specific), a plastidial oleate desaturase (fad6), a plastidial linoleate desaturase (fad7/fad8), a monogalactosyldiacyl-glycerol desaturase (palmite-specific), a digalactosyldiacyl-glycerol synthase, a sulfolipid biosynthesis protein, a long-chain acyl-CoA synthase, an ER glycerol-3-phosphate acyltransferase, an ER 1-acyl-sn-glycerol-3-phosphate acyltransferase, an ER phosphatidic acid phosphatase, a diacylglycerol cholinephosphotransferase, an ER oleate desaturase (fad2), an ER linoleate desaturase (fad3), an ER cytidine-5'-diphosphate-diacylglycerol synthase, an ER phosphatidylglycero-phosphate synthase, an ER phosphatidylglycerol-3-phosphate phosphatase, a Phosphatidylinositol synthase, a diacylglycerol kinase, a choline cholinephosphate cytidylyltransferase, a phosphatidylethanol transfer protein, a choline kinase, a Lipase, a phospholipase C, a phospholipase D, a phosphatidylserine decarboxylase, a phosphatidylinositol-3-kinase, a ketoacyl-CoA synthase (KCS), a (beta)-keto-acyl reductase, and a CER2, a fatty acid isomerase; and

the step of screening the library comprises screening for any of a change in a physical property of one or more lipid, fatty acid, wax or oil in the presence of a polypeptide or RNA encoded by the one or more lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, a protein-protein interaction in a two hybrid assay, expression of a reporter gene in a one hybrid assay, growth or survival of a recombinant cell expressing the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in an elevated temperature environment, growth or survival of a recombinant cell expressing the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a medium comprising a membrane active compound, relative bioluminescence of a recombinant cell comprising at least one gene from the lux operon and the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid, detection of cellular localization of a protein

77

encoded by the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid to a chloroplast, or endoplasmic reticulum, and detection of cellular localization of a product produced as a result of expression of the one or more recombinant lipid biosynthetic nucleic acid or the selected shuffled lipid biosynthetic nucleic acid in a cell.

51. The method of claim 42, wherein the library to be screened is present in a population of cells and the library is selected by growing the cells in or on a medium comprising a cell membrane disruptive agent.

52. The method of claim 42, the method further comprising selecting the library for one or more additional lipid biosynthetic activity.

53. The method of claim 42, wherein the step of recombining is performed in a plurality of cells; the method optionally further comprising one or more of:

(a) recombining DNA from the plurality of cells that display a selected lipid biosynthetic phenotype with a second library of DNA fragments, at least one of which undergoes recombination with a segment in a nucleic acid present in the cells to produce recombined modified lipid synthetic cells, or recombining DNA between the plurality of cells that display a selected lipid biosynthetic phenotype to produce modified lipid synthetic cells;

(b) recombining and screening the recombined or modified cells to produce further recombined cells that have evolved additionally distinct or improved lipid synthetic activity; and

(c) repeating (a) or (b) until the further recombined cells have acquired additionally distinct or improved lipid synthetic activity.

54. The method of claim 53, wherein the method further comprises:

(iii) recombining at least one distinct or improved recombinant lipid synthetic activity with a further lipid synthetic activity, which further nucleic acid is the same or different from one or more of the plurality of parental nucleic acid forms of (i), to produce a further library of lipid synthetic nucleic acids;

(iv) screening the further library to identify at least one further distinct or improved lipid synthetic nucleic acid that exhibits a further improvement in lipid

78

synthetic capability compared to a non-recombinant lipid synthetic gene; and, optionally,

(v) repeating (iii) and (iv) until the further distinct or improved recombinant nucleic acid shows additionally distinct or improved lipid synthetic properties.

55. The method of claim 42, wherein the library is present in bacterial, plant, fungal or animal cells and the method comprises:

pooling multiple separate library members;

screening the resulting pooled library members for a distinct or improved

recombinant lipid biosynthetic nucleic acid that exhibits distinct or improved lipid biosynthesis compared to a non-recombinant lipid biosynthetic nucleic acid; and,

cloning the distinct or improved recombinant lipid biosynthetic nucleic acid from the pooled library members.

56. The method of claim 55, further comprising transducing the distinct or improved recombinant lipid biosynthetic nucleic acid into a plant.

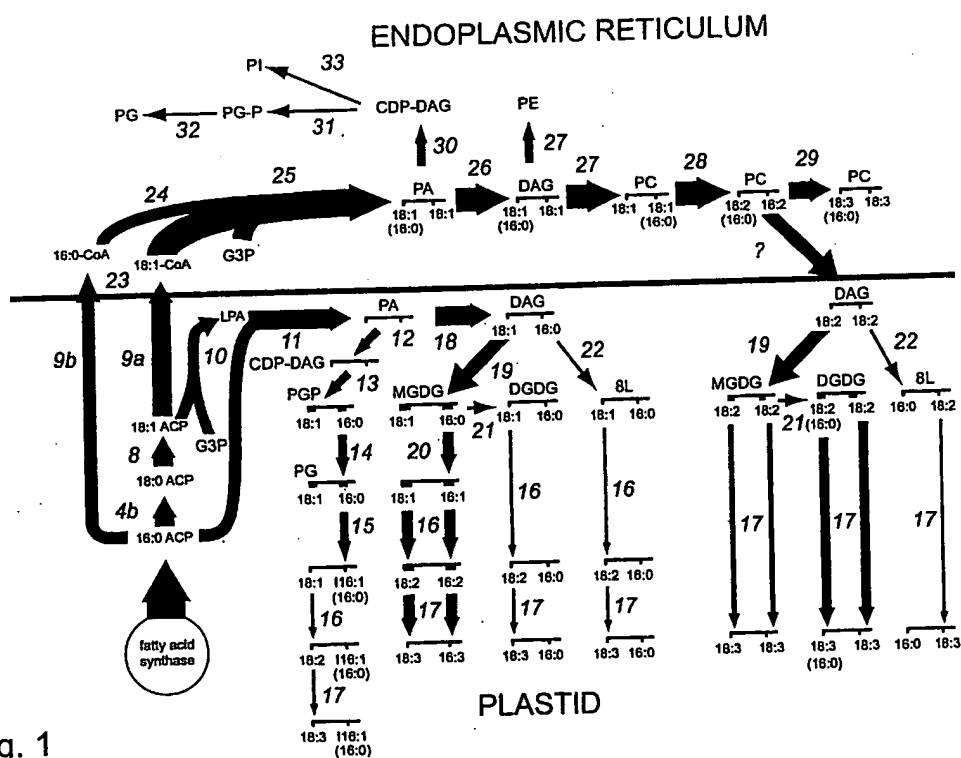
57. The method of claim 56, further comprising transducing the distinct or improved recombinant lipid biosynthetic nucleic acid into a plant and testing the resulting transduced plant for distinct or improved lipid biosynthetic properties.

58. The method of claim 56, further comprising transducing the distinct or improved recombinant lipid biosynthetic nucleic acid into a plant and breeding the plant with a separate plant strain of the same species, followed by selection of resulting offspring for distinct or improved lipid biosynthetic properties.

59. A library of lipid synthetic nucleic acids made by the method of claim 42.

60. The library of claim 58, wherein the library is a phage display library, or wherein the library is present in a bacterial or plant cell.

61. The at least one recombinant lipid biosynthetic nucleic acid made by the method of claim 42.



2/5

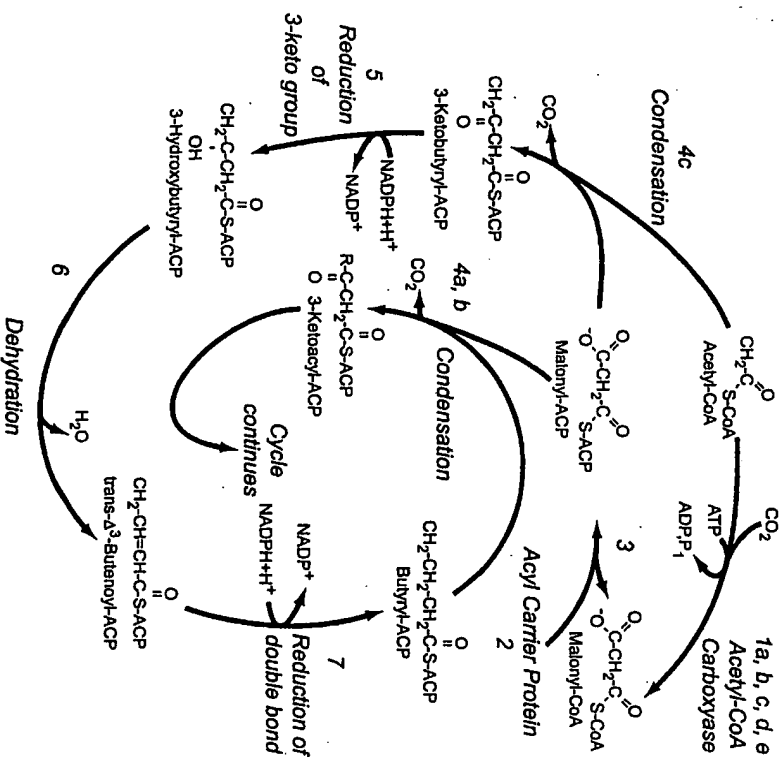


Fig. 2

3/5

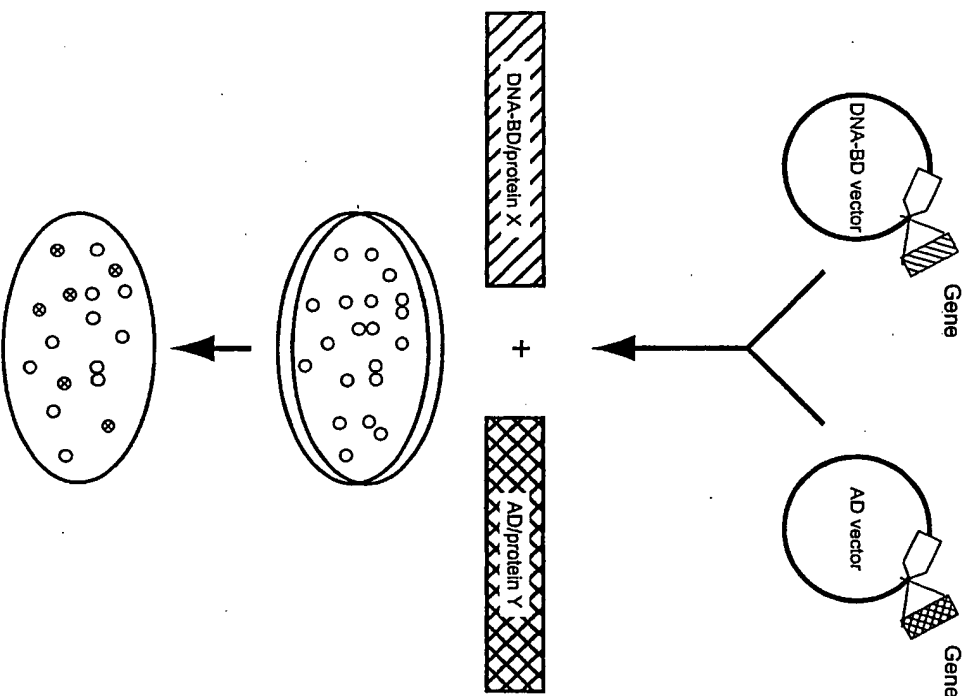


Fig. 3A

4/5

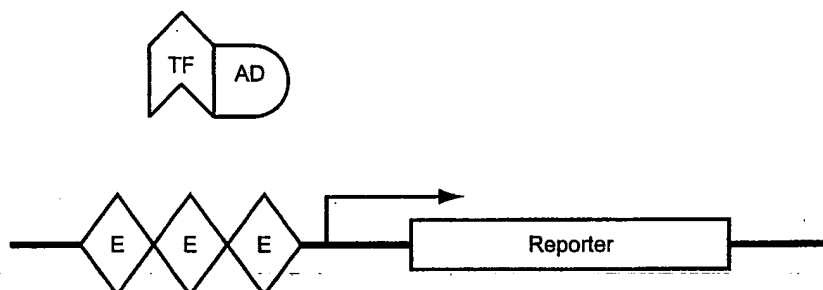
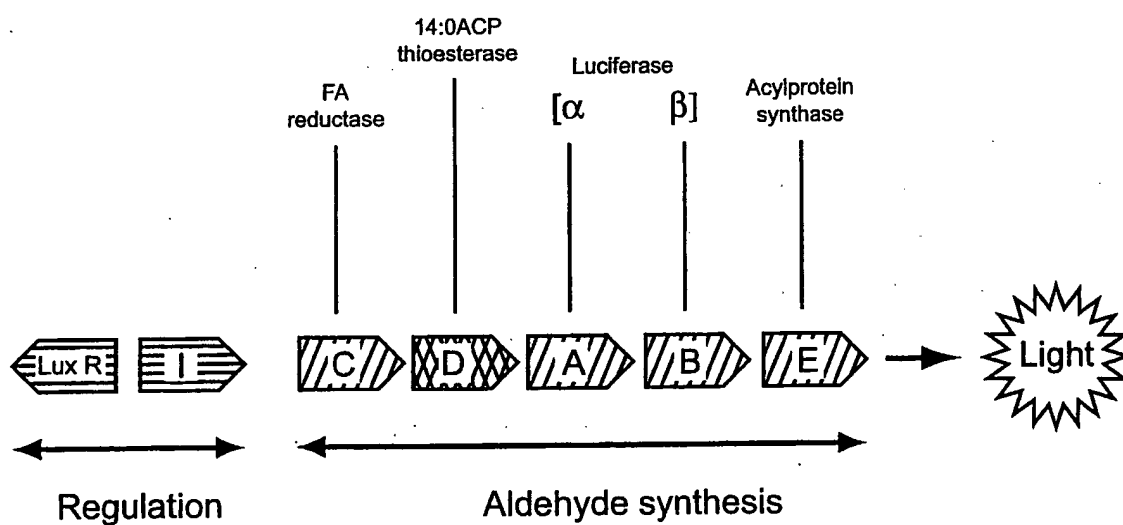


Fig. 3B



5/5

Fig. 4

## INTERNATIONAL SEARCH REPORT

Title International Application No.  
PCT/US 00/09285A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C12M15/10 C12M15/82 A01H5/00B. FIELD(S) SEARCHED  
International documentation searched (classification system followed by classification symbols)  
IPC 7 C12M A01H

Documentation searched other than international documentation to the extent that such documents are included in the data searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)  
EPO-internal, MEDLINE, EMBASE, CHEM ABS Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category \* Citation of document, with indication, where appropriate, of the relevant passages

Relevant to claim No.

X DE 197 31 990 A (STUDIENGESellschaft KOHLE MBH) 28 January 1999 (1999-01-28)

1-4,  
10-12,  
14,  
25-32,  
36,37,61

the whole document

X REEZY M T ET AL.: "Overexpression, immobilization and biotechnological application of Pseudomonas lipases" CHEMISTRY AND PHYSICS OF LIPIDS, IR, LIMERICK, vol. 93, 1 June 1998 (1998-06-01), pages 3-14, XP002087469 ISSN: 0009-3084 see esp. pp.11/12

1-4,  
10-12,  
14,  
25-32,  
36,37,61

-/-

X Further documents are listed in the continuation of box C.

X Patent family members are listed in annex.

Special categories of cited documents:

\* Document defining the general state of the art which is not considered to be of particular relevance

\* Earlier document but published on or after the international filing date

\* Document which may form a basis for a priority claim(s) or which is cited to establish the publication date of another document or other special reason (as specified)

\* Document relating to an oral disclosure, use, exhibition or other means published prior to the international filing date but later than the priority date claimed

\* Document published after the international filing date but not in conflict with the application but which is cited to establish the principle of theory underlying the invention

\* Document of particular relevance; the cited invention cannot be considered novel or cannot be considered to involve an inventive step when the cited invention is taken into account

\* Document of particular relevance; the cited invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents

\* Document which is cited to establish the publication date of another document or other special reason (as specified)

\* Document number of the same patent family

4 August 2000

Date of mailing of the international search report

Name and mailing address of the ISA

European Patent Office, P.O. Box 1  
T 4 (431-70) 540-500, T 3 (31 65) 500 N,  
Fax (43-70) 540-5018

Authorized officer

Kania, T

Form PCT/ISA(21) (second sheet) July 1992

## INTERNATIONAL SEARCH REPORT

Title International Application No.  
PCT/US 00/09285

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category \* Citation of document, with indication, where appropriate, of the relevant passages

Relevant to claim No.

X WO 98 41522 A (NOVONORDISK AS) 24 September 1998 (1998-09-24)

see the whole document; esp. claims 13,14

X FERRI, STEFANO R. ET AL.: "Substrate specificity modification of the stromal glycerol-3-phosphate acyltransferase" ARCH. BIOCHEM. BIOPHYS. (1997), 337(2), 202-208, XP000929718

1-14,  
25-37,61

Y the whole document

15-24,  
38-60

X CAHOON E. ET AL.: "Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position" PNAS U.S.A., vol. 94, 1997, pages 4872-4877, XP002144277

1-14,  
25-37,61

Y the whole document

15-24,  
38-60

Y WO 97 35966 A (MAXYGEN INC.; HINSKILL JEREMY (US); STEINER WILHELM P C (US)) 2 October 1997 (1997-10-02)

see the whole document; esp. p.46ff. part V

15-24,  
38-60

A WO 98 27230 A (MAXYGEN INC.; PATTEN PHILLIP A (US); STEINER WILHELM P C (US)) 25 June 1998 (1998-06-25)

1-61

the whole document

1-61

A BORNISCHER U.T. ET AL.: "Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of opothilones" BIOTECHNOLOGY AND BIOENGINEERING, (5 JUN 1998) 58/5 (354-559), XP002144278

Y the whole document

1-61

A HARAYAMA S.: "Artificial evolution by DNA shuffling" TRENDS IN BIOTECHNOLOGY 68, ELSEVIER PUBLICATIONS, CAMBRIDGE, vol. 16, no. 2, 1 February 1998 (1998-02-01), pages 76-82, XP004107046 ISSN: 0167-7799

see esp. pp.77/78

-/-

Form PCT/ISA(21) (Continuation of second sheet) July 1992

# INTERNATIONAL SEARCH REPORT

Int. Serial Application No.  
PCT/US 00/09285

## C(Contributors) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Character of document, with indication, where appropriate, of the relevant passages	Reference to claim No.
A	CRAMER A ET AL: "DNA shuffling of a family of genes from diverse species accelerates directed evolution." NATURE, (1998 JAN 15) 391 (6664) 288-91. XP000775869 the whole document	1-61
A	STEMMER W: "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, US NATIONAL ACADEMY OF SCIENCE, WASHINGTON, 1994 (1994-10-01), vol. 91, 1 October 1994, XP002087463 pages 10747-10751, XP002087463 ISSN: 0027-8424 cited in the application the whole document	1-61
A	SCHMIDT-DANNERT C ET AL: "Directed evolution of industrial enzymes" TRENDS IN BIOTECHNOLOGY, NL, ELSEVIER, AMSTERDAM, vol. 17, no. 4, Apr'11 1999 (1999-04), pages 135-136, XP004162829 ISSN: 0167-7799 see esp. p.136, 1.col., 5.par.	1-61

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Serial Application No.  
PCT/US 00/09285

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
DE 19731990 A	28-01-1999	WO 9905288 A EP 1002100 A	04-02-1999 24-05-2000
WO 9841622 A	24-09-1998	AU 6611498 A EP 0996718 A	12-10-1998 03-05-2000
WO 9735966 A	02-10-1997	US 5837458 A AU 2337797 A AU 2542697 A CA 2247930 A EP 0932670 A EP 0906418 A WO 9735957 A AU 713952 B AU 1087397 A CA 2239099 A EP 0876509 A EP 0911396 A JP 2000500981 T WO 9720078 A	17-11-1998 17-10-1997 17-10-1997 02-10-1997 04-08-1999 07-04-1999 02-10-1997 16-12-1999 19-06-1997 05-06-1997 11-11-1998 28-04-1999 02-02-2000 05-06-1997
WO 9827230 A	25-06-1998	AU 5729298 A EP 0946755 A	15-07-1998 06-10-1999

**THIS PAGE BLANK (USPTO)**

## European Patent Register Extract

AN: 00921852.0 PN: // ( )  
EUROPEAN PATENT REGISTER / EPIDOS  
04/09/2002

## CHAPTER - PCT DATA BEFORE PUBLICATION BY THE EPO

APPLICATION NUMBER : 00921852.0  
INTERNATIONAL SEARCH REPORT : 19.10.2000  
INTERNATIONAL APPLICATION NR. : US0009285  
INTERNAT. PUBL. NUMBER AND DATE : WO00061740 19.10.2000  
FILING DATE : 06.04.2000  
PRIORITY : 10.04.1999/ USP 128707  
CLASSIFICATION : C12N15/10, C12N15/82, A01H5/00  
DESIGNATION : AT BE CH CY DE DK ES FI FR GB GR IE IT  
TITLE : LI LU MC NL PT SE  
APPLICANT : MODIFIED LIPID PRODUCTION  
: FOR : ALL DESIGNATED STATES  
MAXYGEN, INC.  
515, GALVESTON DRIVE  
REDWOOD CITY, CA 94063/US  
INVENTOR : YUAN, LING/44228 COUNTRY CLON DRIVE/EL  
MACERO, CA 95618/US  
RAILLARD, SUN, AI/964 TROPHY DRIVE/  
MOUNTAIN VIEW, CA 94043/US  
LASSNER, MICHAEL/891 BALBOA LENEUE/  
FOSTERCITY, CA 94404/US

## PCT - CHAPTER II

REQUEST FOR PRELIMINARY EXAMINATION FILED: 19.10.2000  
COUNTRY/DATE DISPATCH/TIME LIMIT/PAYMENT DATE : AT/21.12.2001/M01/00.00.0000  
BE/21.12.2001/M01/00.00.0000  
CH/21.12.2001/M01/00.00.0000  
CY/21.12.2001/M01/00.00.0000  
DE/21.12.2001/M01/00.00.0000  
DK/21.12.2001/M01/00.00.0000  
ES/21.12.2001/M01/00.00.0000  
FI/21.12.2001/M01/00.00.0000  
FR/21.12.2001/M01/00.00.0000  
GB/21.12.2001/M01/00.00.0000  
GR/21.12.2001/M01/00.00.0000  
IE/21.12.2001/M01/00.00.0000  
IT/21.12.2001/M01/00.00.0000  
LU/21.12.2001/M01/00.00.0000  
MC/21.12.2001/M01/00.00.0000  
NL/21.12.2001/M01/00.00.0000  
PT/21.12.2001/M01/00.00.0000  
SE/21.12.2001/M01/00.00.0000  
COMMUNICATION R85B EXAMINATION FEE : 21.12.2001/M01/00.00.0000  
DATE DISPATCH/TIME LIMIT/PAYMENT DATE : 21.12.2001/M01/00.00.0000  
COMMUNICATION R85A NATIONAL BASIC FEE : 21.12.2001/M01/00.00.0000  
DATE DISPATCH/TIME LIMIT/PAYMENT DATE : 21.12.2001/M01/00.00.0000  
DEEMED WITHDRAWN : 13.11.2001  
CHAPTER - APPLICATION DEEMED TO BE WITHDRAWN

## COMMUNICATION,

THAT THE APPLICATION IS DEEMED TO BE WITHDRAWN  
DATE DISPATCH/LEGAL EFFECT DATE : 11.03.2002/13.11.2001  
REASON: A.90(3)/A.78(2)/R.15(2)/R.25

**THIS PAGE BLANK (USPTO)**

European Patent Register Extract

CHAPTER - RENEWAL FEES (ART.86)

-----  
ADDITIONAL FEE A.86(2)  
YEAR/DUE DATE/TIME LIMIT/PAYMENT DATE : 03/30.04.2002/M06/00.00.0000  
\*\* END OF DATA \*\*

**THIS PAGE BLANK (USPTO)**